



Centre de recherche sur les transports (C.R.T.)
Centre for Research on Transportation

CRT-2004-25

TRANSPORTATION ON DEMAND

by

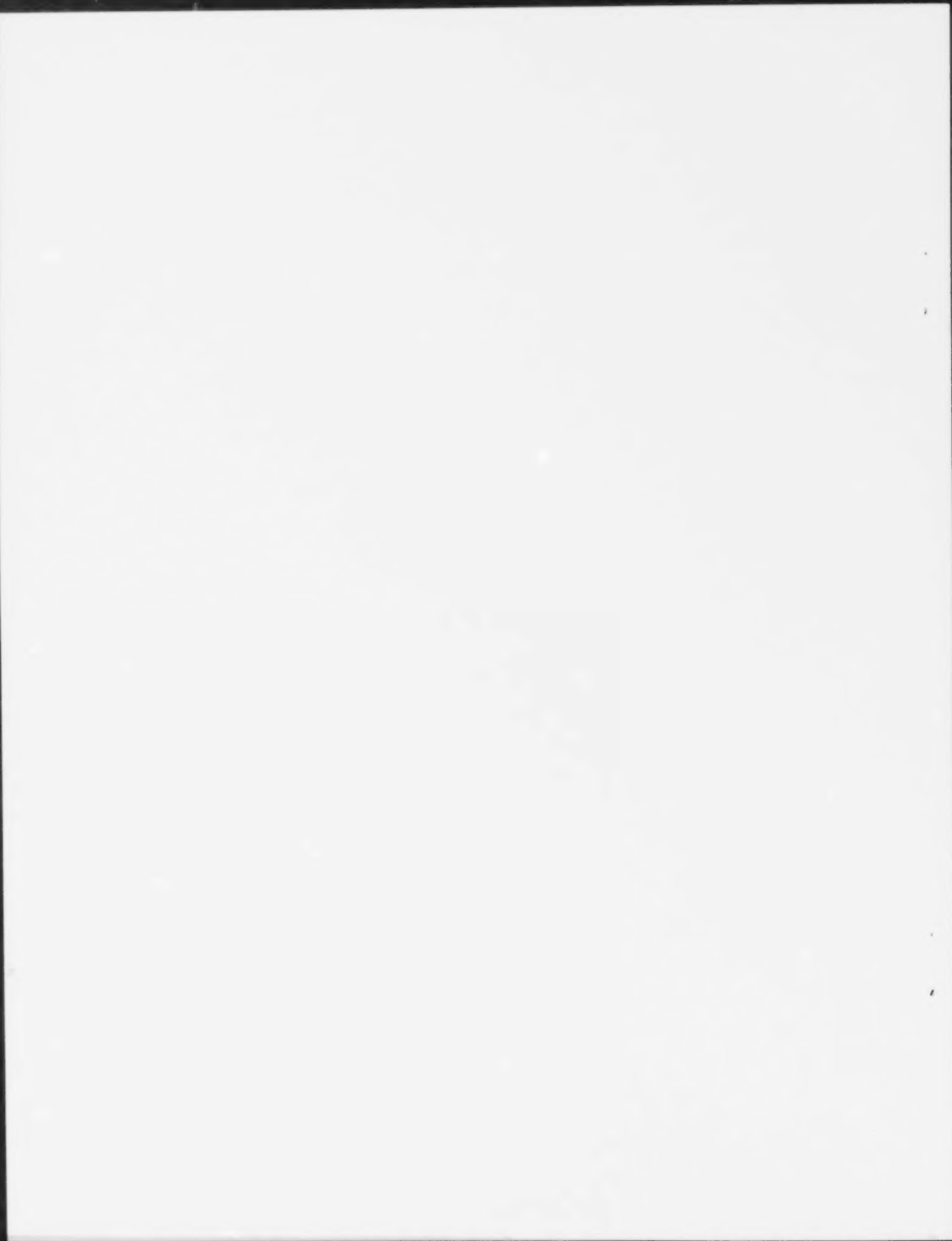
**Jean-François Cordeau
Gilbert Laporte
Jean-Yves Potvin
Martin W.P. Savelsbergh**

October 2004

Université de Montréal
C.P. 6128, succursale Centre-ville, Montréal QC H3C 3J7 Canada
Téléphone: (514) 343-7575 / Télécopieur: (514) 343-7121

crt@crt.umontreal.ca
www.crt.umontreal.ca





**PUBLICATION
CRT-2004-25**

TRANSPORTATION ON DEMAND

by

**Jean-François Cordeau¹
Gilbert Laporte¹
Jean-Yves Potvin²
Martin W.P. Savelsbergh³**

This research was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC).

- ¹ Centre de recherche sur les transports, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, Canada H3C 3J7, and Canada Research Chair in Distribution Management, HEC Montréal, 3000, Côte-Sainte-Catherine, Montréal, Canada H3T 2A7
- ² Département d'informatique et de recherche opérationnelle, Université de Montréal, and Centre de recherche sur les transports, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, Canada H3C 3J7
- ³ School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205, U.S.A.

Dépôt légal – Bibliothèque nationale du Québec
Bibliothèque nationale du Canada, 2004

Abstract

This paper describes the vehicle routing problem with pickup and delivery and reviews the literature on four types of transportation on demand problems: dial-a-ride, urban courier services, dial-a-flight and ambulance fleet management.

Keywords: pickup and delivery, transportation on demand, dial-a-ride, dial-a-flight, ambulance.

Résumé

Cet article décrit le problème de tournées de véhicules avec cueillette et livraison et présente une revue de la littérature pour quatre problèmes de transport à la demande : le transport de passagers, le transport de colis en milieu urbain, le partage d'avions et la gestion de flottes d'ambulance.

Mots-clés : cueillette et livraison, transport à la demande, passagers, avions, ambulances.



1 Introduction

Transportation on demand (TOD) is concerned with the transportation of passengers or goods between specific origins and destinations at the request of users. Common examples are dial-a-ride transportation services for the elderly and the disabled, urban courier services, aircraft sharing, and emergency vehicle dispatching. In all such systems, users formulate requests for transportation from a pickup point to a delivery (or drop-off) point. These requests are served by a set of capacitated vehicles that often provide a shared service in the sense that several passengers or goods may be in a vehicle at the same time.

In recent years, TOD systems have become increasingly popular for a number of reasons. With the ageing of the population and the trend toward the development of ambulatory health care services, more and more people rely on door-to-door transportation systems provided by local authorities. Aircraft sharing has also gained in popularity thanks to cost reduction efforts made by organizations and to the numerous problems that have recently plagued the airline industry. Finally, a growing emphasis on electronic commerce, cycle-time compression and just-in-time deliveries has increased the need for demand-responsive freight transportation systems.

TOD systems can be either static or dynamic. In the first case, all requests are known beforehand while in the second case requests are received dynamically and vehicle routes must be adjusted in real-time to meet demand. For instance, courier services are generally highly dynamic whereas dial-a-ride systems can be regarded as mostly static since they usually require users to make a reservation at least one day in advance. In practice, dynamic problems are often treated as sequences of static subproblems. Reoptimization from the current solution can be performed whenever a new request is formulated, or requests can be buffered and periodically incorporated in the existing vehicle routes in batches.

Most TOD problems are characterized by the presence of three often conflicting objectives: maximizing the number of requests served, minimizing operating costs and minimizing user inconvenience. A balance between these objectives is sometimes obtained by first maximizing the number of requests that can be accepted given the available capacity and then minimizing the operating costs while imposing service quality constraints. Service quality is usually measured in terms of deviations from desired pickup and delivery times and, in the case of passenger transportation, in terms of excess ride time (i.e., the difference between the actual ride time of a user and the minimum possible ride time). Operating costs are mostly related to the number of vehicles used, to total route duration and to total distance traveled by the vehicles.

Another distinguishing aspect of TOD problems is the importance of the temporal dimension. Pickups and deliveries are often restricted to take place within specified time windows. These time windows are sometimes very narrow, especially in the case of passenger transportation. In this context, quality of service is also often controlled by imposing a limit on the ride time of each user. The latter is particularly important in the case of emergency vehicles. Waiting while passengers are in a vehicle can also be prohibited. Finally, maximum route duration constraints are sometimes imposed to take driver shift lengths into account.

The day-to-day management of a TOD system involves making decisions regarding three main aspects: request clustering, vehicle routing and vehicle scheduling. Request clustering consists of creating groups of requests to be served by the same vehicle because of their spatial and temporal proximity. Given these groups, vehicle routing consists of deciding the order in which the associated pickup and delivery locations should be visited by each vehicle. Finally, vehicle scheduling specifies the exact time at which each location should be visited. These decisions are obviously tightly intertwined and a proper management of the system calls for their simultaneous optimization.

The Operations Research literature contains numerous studies addressing both static and dynamic TOD problems. Most variants are in fact generalizations of the Vehicle Routing Problem with Pickup and Delivery (VRPPD). The aim of this chapter is to present the most important results regarding the VRPPD and to survey four areas of applications: the dial-a-ride problem, the urban courier service problem, the dial-a-flight problem, and the emergency vehicle dispatch problem.

The remainder of the chapter is organized as follows. The next section formally defines the VRPPD, introduces notation that will be used throughout the chapter and reviews the related literature. The following four sections then each focus on a specific application by describing the particularities of the problem and summarizing the main exact and heuristic solution algorithms that have been proposed in the literature.

2 The Vehicle Routing Problem with Pickup and Delivery

The VRPPD is a generalization of the classical VRP which also belongs to a larger family of pickup and delivery problems (PDPs). One can distinguish between three well-known types of pickup and delivery problems that have been studied in the literature. One is the single-commodity PDP in which a single type of goods is either picked up or delivered at each node (see, e.g., HERNÁNDEZ-PÉREZ and SALAZAR-GONZÁLEZ, 2004). This is the case, for example, when an armoured vehicle transports money between the branch offices of a bank. Another variant is the two-commodity PDP where two types of goods are considered and each node may act as both a pickup and a delivery node (see, e.g., GENDREAU et al., 1999; ANGELELLI and MANSINI, 2001). This problem arises, for instance, in beer or soft drinks delivery where vehicles deliver full bottles and collect empty ones. A variant of this problem is the VRP with backhauls in which all deliveries must be performed before any pickup. Finally, the n -commodity problem occurs when each commodity is associated with a single pickup node and a single delivery node. This is the case when passengers or goods must be transported from an origin to a destination. This problem is usually referred to as the VRPPD.

Because most practical applications of the VRPPD include restrictions on the time at which each location may be visited by a vehicle, it is convenient to present a slightly more general variant of the problem, called the VRPPD with time windows (VRPPDTW).

Let n denote the number of requests to be satisfied. Assuming that all vehicles are based at a single depot, the VRPPDTW may be defined on a directed graph $G = (N, A)$ where $N = P \cup D \cup \{0, 2n + 1\}$, $P = \{1, \dots, n\}$ and $D = \{n + 1, \dots, 2n\}$. Subsets P and D contain pickup and delivery nodes, respectively, while nodes 0 and $2n + 1$ represent the origin and destination depots. With each request i are thus associated an origin node i and a destination node $n + i$. Let K be the set of vehicles and let $m = |K|$. Each vehicle $k \in K$ has a capacity Q_k and the total duration of its route cannot exceed T_k . With each node $i \in N$ are associated a load q_i and a non-negative service duration d_i such that $q_0 = q_{2n+1} = 0$, $q_i = -q_{n+i}$ ($i = 1, \dots, n$) and $d_0 = d_{2n+1} = 0$. A time window $[e_i, l_i]$ is also associated with each node $i \in N$, where e_i and l_i represent the earliest and latest time, respectively, at which service may begin at node i . With each arc $(i, j) \in A$ are associated a routing cost c_{ij} and a travel time t_{ij} .

For each arc $(i, j) \in A$ and each vehicle $k \in K$, let $x_{ij}^k = 1$ if and only if vehicle k travels from node i to node j . For each node $i \in N$ and each vehicle $k \in K$, let B_i^k be the time at which vehicle k begins service at node i , and Q_i^k be the load of vehicle k after visiting node i . The VRPPDTW can be formulated as the following mixed-integer program:

$$\text{Minimize } \sum_{k \in K} \sum_{i \in N} \sum_{j \in N} c_{ij}^k x_{ij}^k \quad (1)$$

subject to

$$\sum_{k \in K} \sum_{j \in N} x_{ij}^k = 1 \quad (i \in P) \quad (2)$$

$$\sum_{j \in N} x_{ij}^k - \sum_{j \in N} x_{n+i,j}^k = 0 \quad (i \in P, k \in K) \quad (3)$$

$$\sum_{j \in N} x_{0j}^k = 1 \quad (k \in K) \quad (4)$$

$$\sum_{j \in N} x_{ji}^k - \sum_{j \in N} x_{ij}^k = 0 \quad (i \in P \cup D, k \in K) \quad (5)$$

$$\sum_{i \in N} x_{i,2n+1}^k = 1 \quad (k \in K) \quad (6)$$

$$B_j^k \geq (B_i^k + d_i + t_{ij})x_{ij}^k \quad (i \in N, j \in N, k \in K) \quad (7)$$

$$Q_j^k \geq (Q_i^k + q_j)x_{ij}^k \quad (i \in N, j \in N, k \in K) \quad (8)$$

$$B_i^k + d_i + t_{i,n+i} \leq B_{n+i}^k \quad (i \in P, k \in K) \quad (9)$$

$$B_{2n+1}^k - B_0^k \leq T_k \quad (k \in K) \quad (10)$$

$$e_i \leq B_i^k \leq l_i \quad (i \in N, k \in K) \quad (11)$$

$$\max\{0, q_i\} \leq Q_i^k \leq \min\{Q_k, Q_k + q_i\} \quad (i \in N, k \in K) \quad (12)$$

$$x_{ij}^k \in \{0, 1\} \quad (i \in N, j \in N, k \in K). \quad (13)$$

The objective function minimizes the total routing cost. Constraints (2) and (3) ensure that each request is served exactly once and that the associated pickup and delivery nodes are

visited by the same vehicle. Constraints (4)-(6) guarantee that the route of each vehicle k starts at the origin depot and ends at the destination depot. The consistence of time and load variables is ensured by constraints (7) and (8). Constraints (9) force the vehicles to visit the pickup node of a request before its delivery node. Finally, inequalities (10) bound the duration of each route while (11) and (12) impose time windows and capacity constraints, respectively.

The VRPPDTW is NP-hard since it generalizes the Traveling Salesman Problem (TSP) known to be NP-hard (GAREY and JOHNSON, 1979). In the presence of time windows, even finding a feasible solution to the problem is NP-hard since the feasibility problem for the TSP with time windows is itself NP-complete (SAVELSBERGH, 1985).

SAVELSBERGH and SOL (1995) considered a slightly more general formulation of the pickup and delivery problem and reviewed the relevant literature on the problem. A more recent survey on pickup and delivery problems was also prepared by DESAULNIERS et al. (2002). In the remainder of this section, we review the most important exact and heuristic solution algorithms for the VRPPD with and without time windows. We first present algorithms for the single-vehicle case, followed by the multiple-vehicle case.

2.1 The single-vehicle VRPPD

The single-vehicle VRPPD is obtained when $m = 1$ in formulation (1)-(13). Although few real-life applications exist for this problem, it may appear as a subproblem in algorithms for the multiple-vehicle case. It is worth mentioning that unlike the single-vehicle VRP which necessarily reduces to an (uncapacitated) TSP, the single-vehicle VRPPD may incorporate capacity constraints. Indeed, because both pickup and delivery nodes are considered, any number of requests may be served by a single vehicle provided that $q_i \leq Q$ for every request.

2.1.1 Exact algorithms

KALANTARI et al. (1985) have presented branch-and-bound algorithms for the single-vehicle case with finite and infinite vehicle capacity. These algorithms, which are modifications of the algorithm of LITTLE et al. (1963) for the TSP, work by eliminating in each branch of the search tree all arcs that would lead to a violation of a precedence constraint. FISCHETTI and TOTH (1989) have developed an additive bounding procedure for the more general TSP with precedence constraints in which some nodes may have one or several predecessors. This procedure combines the lower bounds obtained from the assignment problem and shortest spanning 1-arborescence problem relaxations, variable decomposition and disjunctions. A lower bounding procedure and a dynamic programming algorithm were later developed by BIANCO et al. (1994) while BALAS et al. (1995) proposed valid inequalities for this problem.

More recently, RULAND and RODIN (1997) introduced a branch-and-cut algorithm for the TSP with Pickup and Delivery (TSPPD). Using the previously introduced notation, the problem can be formulated on an undirected graph $G' = (N, E)$ with binary edge variables x_e , $e \in E$.

Given a node set $S \subseteq N$, denote by $E(S)$ and $\delta(S)$ the sets of edges with both endpoints in S , and with one endpoint in S and the other in $N \setminus S$, respectively. Let also $x(S) = \sum_{e \in E(S)} x_e$. Similarly, let $x(E') = \sum_{e \in E'} x_e$ for any edge set $E' \subseteq E$. Finally, let $\mathcal{U}_s = \{U \subset N \mid 0 \in U\}$ and $\mathcal{U}_p = \{U \subset N \mid 0 \in U, 2n+1 \notin U, \exists i \in P, i \notin U, n+i \in U\}$. The formulation can be stated as follows:

$$\text{Minimize } \sum_{e \in E} c_e x_e \quad (14)$$

subject to

$$x(\{0, 2n+1\}) = 1 \quad (15)$$

$$x(\delta(\{i\})) = 2 \quad (i \in N) \quad (16)$$

$$x(\delta(U)) \geq 2 \quad (U \in \mathcal{U}_s) \quad (17)$$

$$x(\delta(U)) \geq 4 \quad (U \in \mathcal{U}_p) \quad (18)$$

$$0 \leq x_e \leq 1 \quad (e \in E) \quad (19)$$

$$x \in \mathbb{Z}^{|E|}. \quad (20)$$

Constraint (15) simply connects the origin depot to the destination depot and ensures that the solution is a Hamiltonian cycle. Each node is then required to have a degree of 2 by constraints (16) while constraints (17) ensure the biconnectedness of the solution. Finally, constraints (18) force the pickup node of each request to be visited before its delivery node.

As explained by RULAND (1995) a careful analysis of constraints (17) and (18) reveals that the cardinality of the sets $\mathcal{U}_s$ and $\mathcal{U}_p$ can in fact be reduced by exploiting the redundancy of some of the associated constraints. The author also shows that the resulting subtour elimination constraints and precedence constraints define faces of the TSPPD polytope.

Two other classes of inequalities were introduced by RULAND (1995). Let $U_1, \dots, U_m \subset N$ be mutually disjoint subsets and let $i_1, \dots, i_m \in P$ be requests such that $0, 2n+1 \notin U_l$ and $i_l, n+i_{l+1} \in U_l$ for $l = 1, \dots, m$ (where $i_{m+1} = i_1$). The following inequality, called a *generalized order constraint*, defines a proper face of the TSPPD polytope:

$$\sum_{l=1}^m x(U_l) \leq \sum_{l=1}^m |U_l| - m - 1. \quad (21)$$

(Note that similar inequalities were also proposed by BALAS et al. (1995) for the precedence-constrained asymmetric TSP.)

Consider two nodes $i, j \in P$ and a subset H such that $\{i, j\} \subseteq H \subseteq N \setminus \{0, n+i, n+j, 2n+1\}$. The following inequality, called an *order matching constraint*, also defines a proper face of the TSPPD polytope:

$$x(H) + x(\{i, n+i\}) + x(\{j, n+j\}) \leq |H|. \quad (22)$$

The latter inequality can in fact be lifted by considering all requests p for which $p \in H$ and $n+p \in N \setminus H$. For each type of inequality, RULAND (1995) describes separation algorithms

relying on the solution of maximum flow problems. Computational results were reported on instances with $n \leq 15$.

For the single-vehicle VRPPD with time windows and capacity constraints, DESROSIERS et al. (1986) have developed an exact forward dynamic programming algorithm to minimize the total distance traveled. A state (S, i) is defined if there exists a feasible path that starts at the depot 0, visits all nodes in $S \subseteq N$ and ends at node $i \in S$. For each such state, two-dimensional labels are used to keep track of the time and distance traveled. A label can be eliminated if there exists no feasible path that starts at node i and visits all remaining nodes. Computational experiments performed on real-life data with tight time windows have shown that the algorithm could very quickly solve instances with $n \leq 40$.

2.1.2 Heuristics

A probabilistic analysis of a simple construction heuristic for the problem without capacity and time windows was performed by STEIN (1978). This heuristic constructs a solution by concatenating two optimal traveling salesman tours: one through the n origins and one through the n destinations. The author showed that if the $2n$ points are drawn independently from the uniform probability distribution over a subset of the Euclidean plane, then the algorithm has an asymptotic performance bound of 1.06. Later, PSARAFTIS (1983) presented a worst-case analysis of a two-phase construction heuristic. In the first phase, an optimal TSP tour is constructed for the $2n$ points. In the second phase, a solution to the pickup and delivery problem is obtained by traversing the TSP tour clockwise until all points are visited. While doing this, points that have already been visited or that correspond to a destination whose origin has not been visited should be skipped. PSARAFTIS showed that if the minimum spanning tree heuristic of CHRISTOFIDES (1976) is used to construct the TSP tour, the heuristic has a worst-case performance ratio of 3.0. He also reported computational experiments indicating that on realistic size instances the average performance of his heuristic was superior to that of STEIN's heuristic. In a related paper, PSARAFTIS (1983) proposed a local search heuristic that extends the TSP interchange procedure of LIN (1965) to handle precedence constraints. In addition, PSARAFTIS described an approach that identifies the best k -interchange in $O(n^k)$ time. Similar ideas were introduced by SAVELSBERGH (1990) in the more general context of constrained routing problems. Later, HEALY and MOLL (1995) have described a variant of local search for the same problem. Their strategy, called *sacri-ficing*, consists of biasing the search in the direction of solutions with larger neighbourhoods of feasible solutions in the hope of improving the overall quality of the local optima found.

For the single-vehicle problem with time windows, VAN DER BRUGGEN et al. (1993) developed a two-phase local search method based on the variable-depth search of LIN and KERNIGHAN (1973) for the TSP. In the first phase, nodes are first sorted in increasing order of the middle of their time window. The resulting ordering is then modified so as to ensure that the pickup node of each request appears before the delivery node and capacity constraints are satisfied. A solution is then constructed by visiting the nodes in that order. This solution may violate some of the time windows. Iterative improvements are then performed in the hope of obtaining a feasible solution. This solution is also further refined by

applying the same exchange procedures with a different objective.

2.2 The multiple-vehicle VRPPD

2.2.1 Exact algorithms

DUMAS et al. (1991) have proposed a set-partitioning formulation of the problem and an exact column generation algorithm. For vehicle $k \in K$, let Ω^k be the set of feasible routes and let c_r^k be the cost of route r . In addition to traditional flow conservation constraints, each route $r \in \Omega^k$ satisfies time windows, capacity constraints, pairing constraints (i.e., node $i \in P$ is visited iff node $n+i \in D$ is also visited by the route) and precedence constraints. For all $i \in P$ and $r \in \Omega^k$, let a_{ir}^k be a binary constant equal to 1 if request i is served by route r of vehicle k , and 0 otherwise. Finally, define a binary variable y_r^k that takes the value 1 if route r is used for vehicle k , and 0 otherwise. The problem can be stated as follows:

$$\text{Minimize } \sum_{k \in K} \sum_{r \in \Omega^k} c_r^k y_r^k \quad (23)$$

subject to

$$\sum_{k \in K} \sum_{r \in \Omega^k} a_{ir}^k y_r^k = 1 \quad \forall i \in P \quad (24)$$

$$\sum_{r \in \Omega^k} y_r^k = 1 \quad \forall k \in K \quad (25)$$

$$y_r^k \in \{0, 1\} \quad \forall k \in K, r \in \Omega^k. \quad (26)$$

This formulation is solved by a branch-and-bound method in which the linear relaxations are solved by column generation. Columns of negative reduced cost are generated by solving a resource-constrained shortest path problem in which the arc costs are modified to reflect the current values of the dual variables associated with constraints (24) and (25). This problem is solved by a dynamic programming algorithm which is very similar to the one described by DESROSIERS et al. (1986) for the single-vehicle case. In this case, however, not all nodes have to be visited by the vehicle. To obtain integer solutions, branching is performed on additional order variables $O_{ij}, i, j \in P \cup \{0, 2n+1\}$ indicating the sequence in which pickups are performed. These decisions are easily transferred to the subproblem and are handled directly by the dynamic programming algorithm through the introduction of an additional label representing the last pickup node visited. Several arc elimination rules are proposed by the authors to reduce the problem size by taking time windows and pairing constraints into account. For example, arc $(i, n+j)$ can be eliminated if the path $j \rightarrow i \rightarrow n+j \rightarrow n+i$ is infeasible even when setting $B_j = e_j$.

The algorithm was successful in solving two real-life instances with 19 and 30 requests, respectively, as well as randomly generated instances involving up to 55 requests but tight capacity constraints. According to the authors, the algorithm works well when capacity

constraints are restrictive and each route serves a small number of requests (i.e., five or fewer).

A similar approach was developed by SAVELSBERGH and SOL (1998). However, because it was intended to solve large-scale instances, it differs from that of DUMAS et al. (1991) in the following respects: *i*) whenever possible, construction and improvement heuristics are used to solve the pricing subproblem; *ii*) a sophisticated column management mechanism scheme is used to keep the column generation master problem as small as possible; *iii*) columns are selected with a bias toward increasing the likelihood of identifying feasible integer solutions during the solution of the master problem; *iv*) branching decisions are made on additional assignment variables z_i^k representing the fraction of request i that is served by vehicle k ; and *v*) a primal heuristic is used at each node of the search tree to obtain upper bounds. Computational results performed by the authors on instances with $n \leq 50$ show that the proposed approach yields high quality solutions in short computing times even when capacity constraints are not very tight.

Very recently, another column generation method was used by XU et al. (2003) to address a complex pickup and delivery problem encountered in long-haul transportation planning. In their problem, there are multiple carriers and multiple vehicle types available to cover a set of pickup and delivery requests, each of which has multiple pickup time windows and multiple delivery time windows. In addition to the classical vehicle capacity, route duration, pairing and precedence constraints, vehicle routes must satisfy compatibility constraints between the requests, the carriers, and the vehicle types, as well as sequencing constraints requiring the goods to be collected and delivered in a last-in first-out sequence, i.e., the goods picked-up last must be the first to be delivered. Constraints regarding maximum driving time and maximum working time are also taken into account, leading to a complex objective function incorporating fixed costs, mileage costs, waiting costs and layover (driver rest) costs. This problem is solved by means of a column generation approach in which the pricing subproblems are solved by fast heuristics. Instead of embedding the column generation method in a branch-and-bound process, the method reaches an integer solution by applying an IP solver to the restricted set of columns generated in solving the linear relaxation of the problem. Comparisons with lower bounds obtained by solving the LP relaxation of the problem exactly through the use of dynamic programming for the pricing subproblem show that the heuristic approaches are capable of generating near-optimal solutions quickly for randomly generated instances with up to 200 requests. Results are also reported on larger instances involving 500 requests.

2.2.2 Heuristics

A tabu search heuristic for the pickup and delivery problem with time windows was developed by NANRY and BARNES (2000). Solutions that violate time window and vehicle capacity constraints are allowed during the search. These authors have considered three types of move. The first removes a node pair $(i, n + i)$ from its current route and reinserts it in a different route. The second swaps two pairs of nodes between two distinct routes. The last consists of moving a single node within its current route. A hierarchical search mechanism is

used to dynamically alternate between these neighbourhoods according to problem difficulty. Computational results are reported on random instances involving up to 100 requests. A similar tabu search heuristic was also developed by LAU and LIANG (2002).

3 The Dial-a-Ride Problem

The Dial-a-Ride Problem (DARP) is a particular case of the VRPPD arising in contexts where passengers are transported, either in groups or individually, between specified origins and destinations. The most common DARP application arises in door-to-door transportation services for elderly or handicapped people. In this context, users often formulate two requests per day: an *outbound* request from home to a destination, and an *inbound* request for the return trip. The DARP distinguishes itself from the basic VRPPD by its focus on controlling user inconvenience. This usually takes the form of constraints or objective function terms relating to waiting time, ride time (i.e., the time spent by a user in the vehicle) as well as deviations from desired departure and arrival times.

In the remainder of this section, we first discuss the scheduling aspect of the problem and then present the most important exact and heuristic algorithms for the static single-vehicle and multiple-vehicle cases, respectively. This is followed by the dynamic case in the last section. An overview of some of these methods can also be found in the CORDEAU and LAPORTE (2003a) survey.

3.1 Scheduling

Because of the focus on controlling user inconvenience, the scheduling aspect of the problem plays a central role in most applications. As a result, the problem of finding an optimal schedule for a given vehicle route has been studied independently in the literature. In general terms, given a sequence of nodes $i_1, i_2, \dots, i_q$ to be visited, the problem of finding an optimal schedule satisfying time windows can be formulated as:

$$\text{Minimize } \sum_{i=1}^q g_i(B_i) \quad (27)$$

subject to

$$B_i - B_{i+1} \leq -t_{i,i+1} - d_i \quad \forall i = 1, \dots, q-1 \quad (28)$$

$$-B_i \leq -e_i \quad \forall i = 1, \dots, q \quad (29)$$

$$B_i \leq l_i \quad \forall i = 1, \dots, q, \quad (30)$$

where $g_i(B_i)$ is a convex function defined with respect to the time window $[e_i, l_i]$. SEXTON and BODIN (1985a,b) observed that some special cases of this scheduling problem can be seen as a network flow problem and thus solved very efficiently. DUMAS et al. (1989) proposed a dual approach to solve the general problem by performing q unidimensional minimizations. In

the special cases where the inconvenience functions are quadratic or linear, the complexity of the algorithm is $O(q)$. The extension of this methodology to handle time-varying, stochastic travel times was addressed by FU (2002).

On a related topic, HUNSAKER and SAVELSBERGH (2002) have devised a procedure for efficiently testing the feasibility of an insertion in construction or improvement heuristics. They considered a variant of the DARP with time windows, waiting time constraints and ride time constraints, and showed how to check in $O(q)$ time whether the insertion of a given request in a route is feasible. CORDEAU and LAPORTE (2003b) also proposed a procedure, based on the *forward time slack* notion introduced by SAVELSBERGH (1992), to sequentially minimize time window constraint violations, route durations and ride times in the context of local search.

3.2 The static single-vehicle DARP

Early work on the single-vehicle dial-a-ride problem was carried out by PSARAFTIS (1980) who studied the "immediate-request" case in which a list of requests should be served as soon as possible. His model assumes that no time windows are specified by the users. Instead the transporter imposes "maximum position shift" constraints limiting the difference between the position of a request in the calling list and its position in the vehicle route. The objective function aims to minimize the sum of route completion time and customer dissatisfaction. Customer dissatisfaction is itself expressed as a weighted combination of waiting time before pick-up and ride time. The problem is solved using a dynamic programming algorithm in which the state space consists of vectors $(L, k_1, \dots, k_n)$ where L denotes the node being currently visited and k_i denotes the status of request i . The status of request i is either 3 if user i has been dropped off, 2 if the user is still in the vehicle or 1 if the user has yet to be picked up. The complexity of this algorithm is $O(n^2 3^n)$ and only small instances can thus be solved. PSARAFTIS also explains how to handle the dynamic case in which new requests occur dynamically in time but no information on future requests is available. In this context, maximum position shift constraints become essential to prevent a request from being indefinitely deferred. In a later paper, PSARAFTIS (1983) extended his approach to handle time windows on departure and arrival times. The new algorithm has the same complexity as the previous one but uses forward instead of backward recursion.

A heuristic approach based on Benders decomposition was later developed by SEXTON and BODIN (1985a,b) who considered one-sided time windows on delivery. Their algorithm iterates between a routing master problem and a scheduling subproblem. The routing problem relaxes the Benders cuts in the objective function and is solved by a route improvement procedure. The scheduling subproblem is shown to be the dual of a network flow problem that can be solved very quickly. These authors minimize a user inconvenience function made up of the weighted sum of two terms. The first measures the difference between the actual travel time and the direct travel time of a user. The second term is the (positive) difference between desired drop-off time and actual drop-off time, under the assumption that the former is at least as large as the latter, late drop-offs being disallowed. The approach was tested on real-life data sets where the number of users varies between 7 and 20.

3.3 The static multiple-vehicle DARP

One of the first heuristics for the multiple-vehicle DARP was proposed by JAW et al. (1986) who impose windows on the pick-up times of inbound requests and on the drop-off times of outbound requests. A maximum ride time, expressed as a linear function of the direct ride time, is given for each user. In addition, vehicles are not allowed to be idle when carrying passengers. A non-linear objective function combining several types of disutility is used to assess the quality of solutions. The authors have developed an insertion heuristic that selects users in order of earliest feasible pick-up time and gradually inserts them into vehicle routes so as to yield the least possible increase in the objective function. The algorithm was tested on artificial instances involving 250 users and on a real data set with 2617 users and 28 vehicles. This approach was also adapted by ALFA (1986) and applied to a practical case in Winnipeg, Canada.

Several of the heuristics proposed for the multiple-vehicle case are two-phase algorithms in which the first phase creates and selects clusters of users that are then combined into vehicle routes in the second phase. An early approach based on this idea is the interactive optimizer described by CULLEN et al. (1981) for the case of a homogeneous fleet. Another clustering method was proposed by BODIN and SEXTON (1986). Their heuristic creates clusters, applies the single-vehicle algorithm of SEXTON and BODIN (1985a,b) to each cluster and then moves users between clusters so as to reduce total user inconvenience. It was applied to real-life instances involving approximately 85 users each.

DUMAS et al. (1989) later improved upon this methodology by creating so-called "mini-clusters" of users, i.e., groups of users to be served within the same area at approximately the same time. Users in a mini-cluster should be transportable by a single vehicle while respecting constraints on time windows, vehicle capacity, pairing and precedence. The mini-clusters are then optimally combined to form feasible vehicle routes, using column generation. In this phase, a time window is imposed on each cluster to ensure feasibility and columns are generated by solving a constrained shortest-path problem. Finally, each vehicle route is optimized by means of the single-vehicle algorithm of DESROSIERS et al. (1986) and a scheduling step is executed to minimize user inconvenience (DUMAS et al., 1990). Instances with up to 200 users are easily solved, while larger instances require the use of a spatial and temporal decomposition technique. The mini-clustering phase was later improved by DESROSIERS et al. (1991) who described a parallel insertion method that relies on the notion of neighbouring requests. Two requests are said to be neighbours if they satisfy the following conditions:

1. $(e_i \leq e_j \leq l_{n+i})$ or $(e_i \leq l_{j+n} \leq l_{n+i})$ or $(e_j \leq e_i \leq l_{n+i} \leq l_{n+j})$;
2. $(t_{ij} + t_{j,n+i} \leq \alpha t_{i,n+i})$ or $(t_{ji} + t_{i,n+j} \leq \alpha t_{j,n+j})$ with $\alpha > 1$;
3. $|\theta_i - \theta_j| \leq \beta$ where θ_i is the angle between a reference axis and the direction from i to $n+i$;
4. $s(i, j) \geq \gamma$ where $s(i, j)$ are the savings in distance obtained by clustering requests i and j together.

Clusters are then constructed in parallel by treating the requests in decreasing order of the direct duration $t_{i,n+1}$, and considering, at each iteration, the creation of a new cluster or the insertion of a request in clusters that contain at least one neighbouring request. Finally, IOACHIM et al. (1995) showed that there is an advantage in terms of solution quality to use an optimization technique for the construction of the clusters. Results were reported on data sets comprising more than 2500 users.

Another study, by BORNDÖRFER et al. (1997), also uses a two-phase approach in which clusters of users are first constructed and then grouped together to form feasible vehicle routes. A cluster is defined as a "maximal subtour such that the vehicle is never empty". In the first phase, a large set of good clusters are constructed and a set partitioning problem is then solved to select a subset of clusters serving each user exactly once. In the second phase, feasible routes are enumerated by combining clusters and a second set partitioning problem is solved to select the best set of routes covering each cluster exactly once. Both set partitioning problems are solved by a branch-and-cut algorithm. On real-life instances, the algorithm cannot always be run to completion and terminates with the best known solution. It was applied to several instances provided by an operator in Berlin and including between 859 and 1771 transportation requests per day.

In another real-life application, TOTH and VIGO (1996) considered a problem in which users specify requests with a time window on their origin or destination. An upper bound proportional to direct distance is imposed on the ride time. Transportation is supplied by a fleet of capacitated minibuses and by the occasional use of taxis. The objective is to minimize the total cost of service. The authors have developed a heuristic consisting of first assigning requests to routes by means of a parallel insertion procedure, and then performing intra-route and inter-route exchanges. Tests performed on instances involving between 276 and 312 requests show significant improvements with respect to the previous hand-made solutions. Further improvements were later obtained by TOTH and VIGO (1997) through the execution of a tabu thresholding post-optimization phase after the parallel insertion step.

More recently, CORDEAU and LAPORTE (2003b) have developed a tabu search heuristic for the problem in which users specify a desired arrival time for their outbound trip and a desired departure time for their inbound trip, and a maximum ride time is associated with each user. Capacity and maximum route duration constraints are also imposed on the vehicles. The search algorithm is based on a simple mechanism that iteratively removes a request from its current route and reinserts it into another route. As is common in such contexts (see, e.g., CORDEAU et al., 1997), intermediate infeasible solutions are allowed during the search through the use of a penalized objective function. As explained in the previous section, whenever the cost of an exchange is evaluated, the schedules of the two routes involved in the exchange must be updated so as to measure the impacts on violations of time windows, route duration constraints and ride time constraints. The algorithm was tested on randomly generated instances with up to 144 users and on six data sets with $n = 200$ or $n = 295$ provided by a Danish transporter.

Finally, CORDEAU (2003) developed a branch-and-cut algorithm for the same version of the problem. This algorithm relies on several types of valid inequalities that are either new inequalities for the problem or adaptations of known inequalities for the TSP, the TSPPD

or the VRP. The first four types of inequalities are liftings of subtour elimination constraints for the symmetric and asymmetric TSP. Consider the simple subtour elimination constraint $x(S) \leq |S| - 1$ for $S \subseteq P \cup D$. In the case of the DARP, this inequality can be lifted in two different ways by taking into account the fact that for each user i , node i must be visited before node $n + i$. For any set $S \subseteq P \cup D$, let $\pi(S) = \{i \in P | n + i \in S\}$ and $\sigma(S) = \{n + i \in D | i \in S\}$ denote the sets of *predecessors* and *successors* of S , respectively. BALAS et al. (1995) have proposed two families of inequalities for the precedence-constrained asymmetric TSP that also apply to the DARP by observing that each node $i \in P \cup D$ is either the predecessor or the successor of exactly one other node. For $S \subseteq P \cup D$, the following inequality, called a successor inequality (or σ -inequality) is valid for the DARP:

$$x(S) + \sum_{i \in S \cap \sigma(S)} \sum_{j \in S} x_{ij} + \sum_{i \in S \setminus \sigma(S)} \sum_{j \in S \cap \sigma(S)} x_{ij} \leq |S| - 1. \quad (31)$$

Similarly, for any set $S \subseteq P \cup D$, the following predecessor inequality (or π -inequality) is valid for the DARP:

$$x(S) + \sum_{i \in S} \sum_{j \in S \cap \pi(S)} x_{ij} + \sum_{i \in S \cap \pi(S)} \sum_{j \in S \setminus \pi(S)} x_{ij} \leq |S| - 1. \quad (32)$$

Directed subtour elimination constraints proposed by GRÖTSCHEL and PADBERG (1985) for the asymmetric TSP can also be lifted in a similar fashion, yielding two other valid inequalities.

Generalized order constraints, introduced by RULAND (1995) for the TSPPD (see inequalities (21)), can also be lifted in two ways as follows:

$$\sum_{l=1}^m x(U_l) + \sum_{l=2}^{m-1} x_{i_1, i_l} + \sum_{l=3}^m x_{i_1, n+i_l} \leq \sum_{l=1}^m |U_l| - m - 1. \quad (33)$$

$$\sum_{l=1}^m x(U_l) + \sum_{l=2}^{m-2} x_{n+i_1, i_l} + \sum_{l=2}^{m-1} x_{n+i_1, n+i_l} \leq \sum_{l=1}^m |U_l| - m - 1. \quad (34)$$

Finally, ride time constraints may give rise to paths that are infeasible in an integer solution but nonetheless feasible in a fractional solution. Forbidding such paths can be accomplished as follows. For any directed path $\mathcal{P} = \{i, k_1, k_2, \dots, k_p, n + i\}$ such that $t_{i, k_1} + d_{k_1} + t_{k_1, k_2} + d_{k_2} + \dots + t_{k_p, n+i} > L$ the following inequality is valid for the DARP:

$$x_{i, k_1} + \sum_{h=1}^{p-1} x_{k_h, k_{h+1}} + x_{k_p, n+i} \leq p - 1. \quad (35)$$

Heuristic separation algorithms are proposed for each type of valid inequality. In addition, several techniques are described to strengthen the formulation and reduce problem size. In computational experiments, the branch-and-cut algorithm was able to solve instances with up to four vehicles and 32 users.

3.4 The dynamic DARP

While early studies on the DARP were often motivated by dynamic settings, the dynamic DARP has received less attention in the literature than its static counterpart. An approach inspired by the work of JAW et al. (1986) was developed by MADSEN et al. (1995) for a real-life problem involving services to elderly and disabled people in Copenhagen. Users may specify a desired pick-up or drop-off time window, but not both. Vehicles of several types are used to provide service, not all of which are available at all times. In addition, some requests arrive dynamically throughout the day. New requests are inserted in vehicle routes taking into account their difficulty of insertion into an existing route. The algorithm was tested on a 300-customer, 23-vehicle instance and the authors report that it was capable of generating good quality solutions within very short computing times.

At about the same time, DIAL (1995) introduced the concept of an Autonomous Dial-A-Ride Transit (ADART) service based on fully automated command-and-control, order-entry and routing-and-scheduling systems implemented on computers on-board vehicles. Here, the system is fully automated: the only human intervention in the process is the customer requesting service. Furthermore, routing-and-scheduling is not done at some central dispatching centre, but is rather distributed among vehicles through an auction mechanism. ADART puts forward a number of interesting ideas and concepts, but has not yet been implemented in practice.

TEODOROVIC and RADIVOJEVIC (2000) have later studied a generic version of the dynamic dial-a-ride problem using fuzzy logic. Their approach exploits the fact that passengers, dispatcher and drivers have a fuzzy notion of travel times, which can thus be expressed with fuzzy sets and numbers. Through fuzzy arithmetic, calculations about arrival times at customers, waiting times, etc. are performed and the qualitative results are provided to different approximate reasoning algorithms to decide about the assignment and insertion of a new request in a vehicle route.

A software system for demand-responsive passenger services, like variably routed buses, conventional and maxi-taxis, was proposed by HORN (2002). The optimization capabilities of the system are based on least-cost insertions of new requests and periodic reoptimization of the planned routes. The latter is a steepest descent approach using a neighbourhood structure which either moves or swaps customers. A so-called "rank-homing" heuristic is also proposed for governing the relocation of idle vehicles. A set of locations, known as cab-ranks, are specified in advance and the heuristic chooses the cab-rank where the idle vehicle should be dispatched. To take a decision, the heuristic exploits information about future patterns of demand at each cab-rank. The system has only been tested in simulated environments.

Finally, COSLOVICH et al. (2003) have addressed a dial-a-ride problem where people might unexpectedly ask a driver for a trip at a given stop. Clearly, these requests must immediately be accepted or rejected. In order to accommodate them, a neighbourhood of solutions is generated off-line by considering different perturbations to the current planned routes. Using this neighbourhood of solutions, more insertion opportunities are available and can be quickly evaluated when an unexpected customer asks for service. Whenever a new customer

is accepted, both the current solution and the neighbourhood must be updated, but this computationally demanding task can be done while the vehicles are moving from one stop to the next. In that case, the time pressure is much less stringent, when compared with the almost immediate response required by unexpected customers.

4 Urban Courier Service Problems

Every large city has a number of courier companies serving pickup and delivery requests for the transportation of letters and small parcels. These requests occur continuously during the day and only a small fraction of these are known in advance (typically, those requests that have been received the previous day, but too late for immediate service). The distinctive features of *Urban Courier Service Problems* (UCSPs) with regard to the other problems presented in this chapter are their inherent dynamic nature and the absence of capacity constraints, due to the small size of letters and parcels.

In these problems, each request is characterized by a pickup and a delivery location, plus a time window for service. A usually fixed fleet of vehicles is available to service the requests. When a new request occurs, it is dispatched and inserted in a least cost fashion in the planned route of one vehicle. This cost typically relates to the total distance traveled by the vehicles plus a penalty for lateness when the vehicle arrives at a location after its time window's upper bound (a vehicle can arrive before the lower bound but, in that case, it must wait up to the lower bound to start its service). With a fixed size fleet, some requests received during the day may remain unserved. This happens, for example, when drivers must be back at some location before a given deadline at the end of the day, when the upper bounds of the time windows are strictly enforced, or when the incremental cost for servicing a request exceeds a tolerance threshold. In practice, these "unserved" requests will be serviced the next day or will be referred to alternative transportation means, including competitors.

This type of problem has not been studied much in the literature. In particular, we are not aware of any exact methods for solving them and only a few heuristics are reported. Furthermore, the only "dynamic" aspect of the problem that has been considered is the occurrence of new requests, although other aspects are certainly of interest like dynamic travel times. In the following, the main algorithms for the UCSP are reviewed.

The single-vehicle dynamic pickup and delivery problem (without time windows) was analyzed by SWIHART and PAPA-STRAVOU (1999) under different routing policies and demand intensities. This paper can be seen as an extension of the work of BERTSIMAS and VAN RYZIN (1991) on the Dynamic Traveling Repairman Problem (DTRP) with single-point customer requests. As the objective is to minimize the time spent in the system, this work is at the interface of vehicle routing and queuing theory. Dynamic contexts with both unit-capacity and multiple-capacity vehicles are analyzed.

4.1 A Tabu Search Heuristic for the UCSP

A natural heuristic approach for the UCSP is to use a cheapest insertion criterion for new requests. GENDREAU et al. (1998) go beyond this principle and reoptimize the planned routes with a tabu search heuristic. The neighbourhood structure is based on ejection chains (GLOVER, 1996), where the pickup and delivery locations of a request are taken from one route and moved to another route, forcing a request from that route to move to yet another route, and so on. The chain may be of any length and may be cyclic or not. The "best" ejection chain is obtained by solving a constrained shortest path problem. The tabu search heuristic also integrates an adaptive memory (ROCHAT and TAILLARD, 1995), which combines high quality solutions to produce new ones.

Due to real-time requirements, different computational techniques are proposed to speed up the neighbourhood evaluation. In addition, a parallel implementation is developed where a master processor distributes the workload among the slaves that run the tabu search processes. A two-level parallelization scheme is used. First, several tabu search threads run in parallel and communicate through the common adaptive memory: they all feed the memory with new improved solutions and get starting solutions from it. Second, each solution (a set of planned routes) within a search thread is partitioned into subsets of routes, and a different tabu search process is associated with each subset. This is a form of intensification, as it allows the search to focus on restricted parts of the solution, while reducing the computational effort as a side effect.

The tabu search heuristic runs between the occurrence of new events. At each input update, it solves the static problem associated with known requests. Because it does not account for future requests, it can be characterized as a "myopic" problem-solving approach. Two different types of event are considered: the occurrence of new service requests, which are truly dynamic events, and the completion of service at customer locations (which can be determined from the current routes, due to deterministic travel times). When a new request is received, the tabu search processes are stopped and their best solution is sent to the master for possible inclusion in the adaptive memory. The new request is then inserted at least cost in each solution contained in the adaptive memory. Once updated, the memory feeds the tabu search processes with new starting solutions. A similar procedure is applied when service is completed at a given location. In that case, the best solution in the adaptive memory is used to identify the vehicle's next destination, and the other solutions in the memory are updated accordingly.

A simulator was developed to test the algorithm under realistic scenarios with up to 30 requests per hour. The computational results demonstrated the superiority of the tabu search heuristic for handling new requests, when compared with more straightforward approaches, like simple insertion heuristics. It is thus useful to optimize the planned routes with sophisticated procedures, even when the optimization does not account for future requests. Although this is not explicitly mentioned by GENDREAU et al. (1998), the benefits mostly arise from the early portion of the planned route (as opposed to the later portion which is likely to be modified with the arrival of new requests). This observation naturally leads to the work described in the next subsection.

4.2 A Double Horizon Strategy for the UCSP

MITROVIĆ-MINIĆ (2001) and MITROVIĆ-MINIĆ et al. (2004) have proposed a double horizon strategy for solving a variant of the UCSP. In this work, the number of vehicles is a free variable, thus allowing all requests to be serviced. Furthermore, each request must be visited before a deadline, which means that the objective reduces to minimizing the total distance traveled. The authors propose a generalization of the short-term rolling horizon approach (where only requests with a time window sufficiently close to the current time are assigned to planned routes (PSARAFTIS, 1988)). Both a short-term and a long-term planning horizon are considered. The latter is introduced to alleviate the adverse long-term effects of apparently good short-term decisions. Basically, the idea is to associate a different objective with each horizon type. The objective associated with the short term horizon is the true objective (i.e., total distance traveled), while the objective associated with the long-term horizon is aimed at introducing large waiting times in the routes to favour the insertion of future requests. Each objective is optimized with a simplified version of the tabu search heuristic reported in Section 4.1. The computational results obtained on instances generated from data collected from two courier companies, operating in Vancouver, Canada, demonstrate the benefits of the double horizon approach when compared to a single horizon approach. MITROVIĆ-MINIĆ and LAPORTE (2004) have further analyzed different ways of introducing waiting times when scheduling the planned routes. They showed that mixed waiting strategies provide better results. The best approach partitions a planned route into segments made of close locations. Within a segment, the vehicle always departs as soon as possible from its current location; but when it is time to cross a boundary between two segments to travel further, the vehicle waits at its current location for a fraction of the time available up to the latest possible departure time.

4.3 Adaptive methods for the UCSP

The approaches reported in this subsection exploit the knowledge accumulated, over the years, by expert human dispatchers to make good dispatching decisions. They can be considered as learning or adaptive methods. In the work of SHEN et al. (1995), a neural network model learns to assign requests to vehicles by automatically adjusting itself to a sample of decisions previously made by an expert (see RUMELHART et al. (1986) for an introduction to feedforward neural networks and the backpropagation learning algorithm). After training with 140 dispatching scenarios obtained from a small courier company operating in Montreal, and for which expert decisions were known, the network was able to take good dispatching decisions on other sets of previously unseen scenarios.

LECLERC and POTVIN (1997) proposed a linear utility function that integrates the main decision variables considered by expert dispatchers when they make decisions. The decision variables within the utility function are then weighted with a genetic algorithm (HOLLAND, 1992). Basically, different sets of weights "evolve" through genetic mechanisms, with the objective of matching as closely as possible a sample of decisions previously taken by an expert. BENYAHIA and POTVIN (1998) extended this work by evolving nonlinear utility

functions, using a genetic programming framework (KOZA, 1992).

5 The Dial-a-Flight Problem

Almost 3,000 businesses in the United States provide on-demand air charter services (certificated by the FAA as Part 135 on-demand air charter). The majority of companies in the industry are small businesses regulated by the FAA with similar oversight to that given to the large scheduled airlines. The on-demand air charter industry provides a vital transportation link for medical services, important cargo needed to promote commerce, and personal travel supporting the growth of the economy. These companies use smaller aircraft to meet the customized needs of the traveling public for greater flexibility in scheduling and access to almost every airport in the country. Flights are planned according to the customer's schedule, not the operator's. On-demand air charter serves commerce across the country and the world by providing short notice delivery of parts, important documents, supplies and other valuable cargo. On-demand air charter also saves lives, since air ambulances transport critically ill or injured patients to hospitals and trauma centers that can provide the necessary care, and transport vital organs for those requiring transplants. All of these services are contingent upon the ability to respond quickly to the needs of customers.

Recently there has been an increased interest in the passenger service sector of the on-demand air charter industry. This is in part due to the changes taking place in the commercial airline industry. Increased security at airports has resulted in longer waiting times, with the associated frustrations, and thus longer travel times. Furthermore, due to the huge losses suffered by the airlines in recent years (airlines worldwide have lost \$25 billion and more than 400,000 jobs in 2002 and 2003), airlines have cut back and are operating with a reduced schedule, affecting the flexibility of the business traveler, especially when it concerns smaller regional airports. At the same time, technological advances are paving the way for the development of cheaper jet airplanes. For example, the Eclipse 500, a six-seat, single-pilot state-of-the art jet will sell for about \$1 million – about one-quarter of the price of the cheapest business jets made today. As a consequence, the idea of an *air taxi* service, providing efficient, hassle-free, affordable, on-demand air transportation, is no longer just fiction; it is rapidly becoming a reality. In fact, an air taxi already exists today. Since April 2002, SkyTaxi, Inc. (www.skytaxi.com) has been providing on-demand air transportation in the northwestern United States. Some passengers pay a premium to fly direct and by themselves; others receive a discount by agreeing to allow stops to pick up or drop off other passengers.

There are obvious advantages to an air taxi system. More than 1.5 million people board commercial airliners each day. Most fly with hundreds of other passengers on jumbo jet airplanes to and from a limited number of major "hub" airports which are heavily congested and often located many miles from their homes and final destinations. Missed connections and flight delays add to their frustrations. An air-taxi system gives travelers the option of hopping aboard small jets that fly to and from less congested outlying airports, without packed parking lots, long lines at security checkpoints, flight delays, and lost luggage, that

are closer to where they live and where they want to go. While commercial flight service now exists at only about 550 airports in the United States, air taxis will be able to land at 10,000 of the nation's 14,000 public and private runways. And all that at competitive fares.

Even though many characteristics of an air taxi service are similar to those of the common taxi cab service, there are also some fundamental differences. Requests for service are typically placed farther in advance, a day to two days in advance as opposed to minutes to an hour. This gives more time to optimize a flight schedule. Usually, a request for service involves several origin-destination pairs, because business travelers in the end want to return home. Furthermore, the set of possible pickup, drop-off, and transit locations is relatively small and known in advance. An air taxi service operates out of a given set of airports, which implies a flight schedule optimizer can exploit the fixed transportation network structure. Finally, the FAA imposes strict rules on pilot flying and duty hours as well as on aircraft maintenance, which all affect scheduling flexibility.

Consequently, the traditional dial-a-ride problem is insufficient as an abstract representation for studying, analyzing, and developing decision technology for air taxi services. In this section, we introduce the *Dial-a-Flight Problem* (DAFP). As air taxi services are a new phenomenon, there is little or no literature on dial-a-flight problems, so we will focus on problem definition and optimization challenges.

5.1 Problem Definition

A key challenge in setting up an air taxi system is the development of a scheduling engine that takes requests for transportation and schedules planes and pilots in a cost effective way to satisfy these requests. The static and dynamic DAFPs defined below capture the essential characteristics of the scheduling problems encountered by an air taxi service.

5.1.1 The Static Dial-a-Flight Problem

The *Static Dial-a-Flight Problem* (SDAFP) is concerned with the scheduling of a set P of n single passenger requests for air transportation during a single day. A request $i \in P$ specifies an origin airport o_i , an earliest acceptable departure time e_i at o_i , a destination airport d_i , and a latest acceptable arrival time l_i at d_i . A request i results in a revenue of r_i . A fleet F of m identical airplanes with capacity Q and operable by a single pilot is available to provide the requested air transportation. Each airplane $j \in F$ has a home base B_j , is available between E_j and L_j , and returns to its home base at the end of the day. A set P of pilots, stationed at the home bases of the airplanes, are available to fly the airplanes. A pilot departs from the home base at the start of his duty and returns to his home base at the end of his duty. A pilot schedule has to satisfy FAA regulations governing flying hours and duty period, i.e., a pilot cannot fly more than 8 hours in a day and his duty period cannot be more than 14 hours. It takes an airplane t_{uv} time to fly from airport u to v (over a distance d_{uv}) and a cost c_{uv} is incurred when doing so. To ensure acceptable service a passenger itinerary will involve at most two flights, i.e., a single intermediate stop is allowed. The turnaround

time at an airport, i.e., the minimum time between an arrival at an airport and the next departure, is 30 minutes. The objective is to maximize the profit, i.e., revenues minus costs, while satisfying all requests. (Note that because all requests have to be satisfied, in this variant of the problem the revenues are fixed and the objective is to minimize the costs.) A dispatcher has to decide which planes and pilots to use to satisfy the requests and what the plane and pilots itineraries will be, i.e., the flight legs and associated departure times.

Several restrictions in the above problem definition represent business decisions rather than physical limitations. For example, the limit on the duration of a passenger trip from origin to destination is a business rule, set, for example, to twice the direct flying time. This constraint which is similar to the maximum ride time in the DARP reflects a trade-off between scheduling flexibility and customer service. Intuitively, an important factor contributing to profitability for an air taxi service provider is a high plane utilization. This can be accomplished by having an airplane satisfy multiple requests at the same time. Therefore, an air taxi service provider may "encourage" passengers to accept a trip that involves an intermediate stop (to pickup or drop off other passengers). Note that the turnaround time also affects flexibility, and therefore profitability. Another business rule, which has not been made explicit in the problem definition, is whether or not to allow passengers to change planes at an intermediate stop. Again, allowing this will increase the scheduling flexibility, but it may have a negative impact on customer service because of potential delays.

5.1.2 The Dynamic Dial-a-Flight Problem

In the *Dynamic Dial-a-Flight Problem* (DDAFP) the set of requests for air transportation arrives over time and each time a request arrives one must immediately decide whether it is feasible to accept the request given the available resources and the commitments already made. In addition, if it is feasible to accept the request, one may also want to decide whether it is desirable to accept the request, i.e., whether it will increase profit. The latter decision is especially complex as it depends on the requests that will arrive in the future.

The simplest variant of the DDAFP is to construct a schedule of flights for a specific day in the not too distant future. Requests for transportation on that particular day arrive in real-time and are considered up to a certain cut-off time, which precedes the actual execution of the planned schedule. The rule is to accept each request if there is available capacity. In the DDAFP, one needs to accommodate bundles of requests, as customers typically request not one, but two or more flights, and ultimately want to return to their point of origin. Clearly, all these flight requests must be accepted as a bundle, rather than individually. Note that all requests are received prior to the execution of any flight schedule. A more complex variant incorporates "same day travel" service, where requests can arrive during the execution of a flight schedule and have to be incorporated into the schedule.

The real-time, online nature of the booking process is not specific to the air taxi service, but a common feature of many transportation problems. The literature on dynamic and stochastic routing problems is rapidly expanding. For a discussion of many of the issues related to accept/reject decisions in transportation problems, see BENT and HENTENRYCK

(2004) and CAMPBELL and SAVELSBERGH (2004, 2003).

5.2 Algorithms for Dial-a-Flight Problems

As mentioned earlier, little or no literature exists on dial-a-flight problems. Therefore, we present a natural integer programming formulation for SDAFP and we discuss some pertinent issues concerning solution approaches for the DDAFP.

5.2.1 A model for the Static Dial-a-Flight Problem

In this subsection, we present a time-discretized multicommodity network flow model for one variant of the problem. We assume the company providing the air taxi service has decided to operate each of its planes with two pilot shifts per day, a morning shift and an afternoon shift. The plane has to return to its home base to switch pilots some time in the early afternoon. The shift lengths are chosen so that the limit on pilot duty hours is automatically satisfied. Finally, it is assumed that for the expected demand distribution it is unlikely that the limit on the number of flying hours of the pilots is violated, and therefore, pilot constraints are not explicitly incorporated in the model.

The time horizon is discretized into time periods of several minutes. Let T be the set of time periods in the planning horizon and let A be the set of airports. Define the following decision variables:

$$x_{uvt}^i = \begin{cases} 1 & \text{if passenger } i \text{ departs from airport } u \text{ to airport } v \text{ at time } t \\ 0 & \text{otherwise} \end{cases}$$

$$y_{uvt}^j = \begin{cases} 1 & \text{if airplane } j \text{ departs from airport } u \text{ to airport } v \text{ at time } t \\ 0 & \text{otherwise.} \end{cases}$$

For each airplane j , let R_j and S_j define the start and end of the period during which the pilot switch needs to take place.

The mathematical formulation of SDAFP can now be written as follows (where parameters have been converted to their time period equivalents whenever necessary):

$$\text{Minimize } \sum_j \sum_u \sum_{u \neq v} \sum_t c_{uv} y_{uvt}^j$$

subject to

$$\sum_v x_{uvt}^i = 1 \quad (i \in P, u = o_i, t = e_i) \quad (36)$$

$$\sum_v x_{vu(t-t_{vu})}^i = 1 \quad (i \in P, u = d_i, t = l_i) \quad (37)$$

$$x_{uu(t-1)}^i + \sum_{v \neq u} x_{vu(t-t_{vu})}^i - \sum_{v \neq u} x_{uvt}^i - x_{uut}^i = 0 \quad (i \in P, u \in A, t = e_i + 1, \dots, l_i - 1) \quad (38)$$

$$\sum_u \sum_{v \neq u} \sum_t x_{uvt}^i \leq 1 \quad (i \in P) \quad (39)$$

$$\sum_i x_{uvt}^i - \sum_j Q y_{uvt}^j \leq 0 \quad (u \in A, v \in A, u \neq v, t \in T) \quad (40)$$

$$\sum_v y_{uvt}^j = 1 \quad (j \in F, u = B_j, t = E_j) \quad (41)$$

$$\sum_v y_{vu(t-t_{vu})}^j = 1 \quad (j \in F, u = B_j, t = L_j) \quad (42)$$

$$y_{uu(t-1)}^j + \sum_{v \neq u} y_{vu(t-t_{vu})}^j - \sum_{v \neq u} y_{uvt}^j - y_{uut}^j = 0 \quad (j \in F, u \in A, t = E_j + 1, \dots, L_j - 1) \quad (43)$$

$$\sum_v \sum_{t=R_j-t_{vu}}^{t=S_j-t_{vu}} y_{vut}^j \geq 1 \quad (j \in F, u = B_j) \quad (44)$$

Constraints (36) and (37) state that each passenger departs from his origin airport and arrives at his destination airport. Constraints (38) ensure passenger flow conservation at airports. Similar constraints are specified for each airplane, i.e., (41), (42), and (43). Constraints (39) limit the number of intermediate stops for each passenger and constraints (40) enforce the airplane capacity, i.e., the total number of passengers flying on a leg is less than the total capacity of all airplanes flying on that leg. Finally, constraints (44) force each airplane to return to its home base to switch pilots.

This time-discretized multicommodity network flow model becomes large quickly and specialized solution approaches need to be developed to solve even medium size instances, e.g., involving 15 to 30 airports and 5 to 10 airplanes.

5.2.2 Algorithmic issues related to the dynamic dial-a-flight problem

In the DDAFP one has to decide, given a set of already accepted requests, whether an incoming request can be served or not. The amount of time available to make this decision depends on the business rules but most likely there is very little time to do so. Consequently, fast heuristics will have to be part of the decision technology. It is, however, conceivable that when heuristics fail to accept a request quickly, a customer may be given the option of receiving final notification of acceptance or rejection in, say, 30 minutes to allow time for optimization based techniques to try and accommodate the request. Research on the

DDAFP is relatively new and, in the current state of knowledge, there are more questions than answers.

Given a set of already accepted requests and an existing feasible schedule, it makes sense to ask what is the value of that schedule in deciding whether it is feasible to accept a new request. What if simple insertion into any of the existing airplane itineraries is not feasible? Does one try to reoptimize a single airplane itinerary or try to reoptimize several airplane itineraries? One by one or simultaneously? How does one select the airplane itineraries to consider? What if the existing schedule included passenger plane changes? In fact, there is no need to keep just one feasible schedule for the already accepted requests. Should one keep several feasible schedules and try to insert new requests in each of them?

Once an incoming request has been accepted, there may be time available to optimize the schedule of the accepted requests by solving a static dial-a-flight problem. Does that help handle the next request? Should the standard objective function be used or is it better to use an objective function that focuses on the remaining flexibility in the schedule? How does one formally define remaining flexibility?

5.3 The dial-a-flight problem in practice

The DAFP is a complex and challenging optimization problem, but still ignores many relevant practical aspects. For example, in reality the costs of a flight leg is a function of the fuel consumption, which depends on the weight of the plane, the altitude of the flight, and the duration of the flight (shorter flights burn more fuel as most fuel is consumed during take off and landing). Furthermore, fuel prices may differ at various locations, so deciding where to refuel may impact overall costs. Also, taxi service providers will likely offer various service classes, e.g., for a higher price a direct flight will be guaranteed, for an even higher price one can charter an entire airplane plus a pilot for certain period of time.

5.4 Performance metrics

Finally, even though profit is the primary concern of an air taxi service, as it is with any other business, many other measures need to be looked at to judge overall performance. As air taxi services receive requests for transportation over time and may or may not accept each request, it is important to keep track of the rejection rate, i.e., the ratio of the number of requests rejected and the number of requests received. To judge the quality of the flight schedule it is of interest to look at the ratio of relocation time (flying without passengers) and total flying time and the average number of passengers per flight leg (either over all flight legs or only over those flight legs that carry passengers). To measure the level of passenger consolidation one can look at the ratio of the sum of the number of passengers on all flight legs and the number of passengers served. (Suppose there are two requests for a single passenger each with origin-destination pairs (A,B) and (A,C). If the schedule contains the two legs (A,B) and (B,C) with two passengers on leg (A,B) and one passenger on leg (B,C), then the ratio is 1.5, indicating that 50% of the passengers fly with an intermediate

stop.) Customer service related measures include the mean ratio of trip time and direct flying time and the mean difference between latest acceptable arrival time and actual arrival time.

6 Ambulance Fleet Management

Transportation on demand problems also arise in the planning of emergency services associated with fire fighting, police patrols and ambulance fleet planning. Several important location, staffing and dispatching issues lie at the heart of the problems encountered in these three areas but ambulance fleet management is certainly the most relevant to this chapter. Readers interested in the management of fire companies and in police patrols are referred to the work of LARSON (1972), WALKER et al. (1979), SWERSEY (1994), and ADAMS (1997).

One of the key issues arising in ambulance fleet management is to decide where to locate ambulances to provide at all times an adequate population coverage. Over the past thirty-five years, there has been a steady evolution in the development of ambulance location models, as witnessed by the work of MARIANOV and REVELLE (1995) and BROTCORNE et al. (2003). The latter study serves as a basis for this chapter. The first models developed in the 1970s proposed static solutions to a problem that is essentially stochastic and dynamic. Over the years more realism has gradually been introduced into these basic models until the emergence, a few years ago, of a truly dynamic solution procedure.

Most contributions in the field of ambulance location present integer linear programming formulations, but no solution techniques. Only in recent years have algorithms been proposed.

6.1 Two early ambulance location models

Ambulance location models are usually defined on a graph $G = (V \cup W, A)$, where V is a node set representing aggregated demand points, W is a set of potential ambulance location sites, and $A = \{(i, j) \in V \cup W\}$ is an arc set. With each arc (i, j) is associated a travel time t_{ij} . A demand point $i \in V$ is *covered* by site $j \in W$ if and only if $t_{ij} \leq r$, where r is a preset coverage standard. Let $W_i = \{j \in W : t_{ij} \leq r\}$ be the set of location sites covering demand point i .

The *Location Set Covering Model* (LSCM) of TOREGAS et al. (1971) aims to minimize the number of ambulances needed to cover all demand points. It uses binary variables x_j equal

to 1 if and only if an ambulance is located at j :

$$\text{(LSCM)} \quad \text{Minimize} \quad \sum_{j \in W} x_j \quad (45)$$

$$\text{subject to} \quad \sum_{j \in W_i} x_j \geq 1 \quad (i \in V) \quad (46)$$

$$x_j \in \{0, 1\} \quad (j \in W). \quad (47)$$

The *Maximum Covering Location Problem* (MCLP) of CHURCH and REVELLE (1974) works with a fixed number p of ambulances and attempts to cover the largest possible demand $z(p)$. Denote by d_i the demand at node $i \in V$ and let y_i be a binary variable equal to 1 if and only if i is covered by at least one ambulance:

$$\text{(MCLP)} \quad \text{Maximize} \quad z(p) = \sum_{i \in V} d_i y_i \quad (48)$$

$$\text{subject to} \quad \sum_{j \in W_i} x_j \geq y_i \quad (i \in V) \quad (49)$$

$$\sum_{j \in W} x_j = p \quad (50)$$

$$x_j \in \{0, 1\} \quad (j \in W) \quad (51)$$

$$y_i \in \{0, 1\} \quad (i \in V). \quad (52)$$

A good way to combine these two models is to repeatedly solve MCLP with increasing values of p and select a solution offering a good compromise between p and $z(p)$.

6.2 Static models with extra coverage

One major drawback of LSCM and of MCLP is that adequate coverage may no longer exist once an ambulance has been dispatched. This is the reason why models with extra coverage have been introduced. The *Tandem Equipment Allocation Model* (TEAM) of SCHILLING et al. (1979) works with two equipment types A and B , corresponding to advanced life support (ALS) units and basic life support units (BLS) operating with different time standards (see MANDELL, 1998). Let r^A and r^B be the respective coverage standards of A and B , and let $W_i^A = \{j \in W : t_{ij} \leq r^A\}$, $W_i^B = \{j \in W : t_{ij} \leq r^B\}$. Let $x_j^A(x_j^B)$ be a binary variable equal to 1 if and only if a vehicle of type $A(B)$ is located at $j \in W$, and let y_i be a binary variable equal to 1 if and only if $i \in V$ is covered by two types of vehicle.

$$\begin{array}{ll} \text{(TEAM)} & \text{Maximize} \quad \sum_{i \in V} d_i y_i \end{array} \quad (53)$$

$$\begin{array}{ll} \text{subject to} & \sum_{j \in W_i^A} x_j^A \geq y_i \quad (i \in V) \end{array} \quad (54)$$

$$\sum_{j \in W_i^B} x_j^B \geq y_i \quad (i \in V) \quad (55)$$

$$\sum_{j \in W} x_j^A = p^A \quad (56)$$

$$\sum_{j \in W} x_j^B = p^B \quad (57)$$

$$x_j^A \leq x_j^B \quad (j \in W) \quad (58)$$

$$x_j^A, x_j^B \in \{0, 1\} \quad (j \in W) \quad (59)$$

$$y_i \in \{0, 1\} \quad (i \in V). \quad (60)$$

This model is a direct extension of MCLP, with the proviso that constraints (58) impose a hierarchy between the two vehicle types. In the FLEET model of SCHILLING et al. (1979), these constraints are relaxed and the number of potential location sites is limited to a preset value p . Another variant, proposed by DASKIN and STERN (1981), is to use a hierarchical objective to first maximize the number of demand points covered more than once.

In the same spirit, HOGAN and REVELLE (1986) have proposed the *Backup Coverage Problems*, called BACOP1 and BACOP2, in which x_j is the number of ambulances located at $j \in W$, and y_i, u_i are binary variables equal to 1 if and only if $i \in V$ is covered once or at least twice, respectively.

$$\begin{array}{ll} \text{(BACOP1)} & \text{Maximize} \quad \sum_{i \in V} d_i u_i \end{array} \quad (61)$$

$$\begin{array}{ll} \text{subject to} & \sum_{j \in W_i} x_j \geq 1 + u_i \quad (i \in V) \end{array} \quad (62)$$

$$\sum_{j \in W} x_j = p \quad (63)$$

$$u_i \in \{0, 1\} \quad (i \in V) \quad (64)$$

$$x_j \geq 0 \text{ and integer} \quad (j \in W), \quad (65)$$

and

$$\text{(BACOP2)} \quad \text{Maximize} \quad \theta \sum_{i \in V} d_i y_i + (1 - \theta) \sum_{i \in V} d_i u_i \quad (66)$$

$$\text{subject to} \quad \sum_{j \in W_i} x_j \geq y_i + u_i \quad (i \in V) \quad (67)$$

$$u_i \leq y_i \quad (i \in V) \quad (68)$$

$$\sum_{j \in W} x_j = p \quad (69)$$

$$u_i \in \{0, 1\} \quad (i \in V) \quad (70)$$

$$y_i \in \{0, 1\} \quad (i \in V) \quad (71)$$

$$x_j \geq 0 \text{ and integer} \quad (j \in W), \quad (72)$$

where θ is a weight chosen in $[0, 1]$.

The *Double Standard Model* (DSM) of GENDREAU et al. (1997) works with two coverage standards r_1 and r_2 , with $r_1 < r_2$, as specified by the United States Emergency Medical Services Act of 1973. A proportion of α of the demand must be covered within r_1 while the entire demand must be covered within r_2 . In the DSM, the objective is to maximize the demand covered twice within r_1 using p ambulances, at most p_j of which are located at $j \in W$, subject to the double coverage constraints. Let $W_i^1 = \{j \in W : t_{ij} \leq r_1\}$ and $W_i^2 = \{j \in W : t_{ij} \leq r_2\}$. The integer variable x_j denotes the number of ambulances located at $j \in W$ and the binary variable y_i^k is equal to 1 if and only if the demand at node $i \in V$ is covered k times ($k = 1$ or 2) within r_1 . The formulation is then:

$$\text{(DSM)} \quad \text{Maximize} \quad \sum_{i \in V} d_i y_i^2 \quad (73)$$

$$\text{subject to} \quad \sum_{j \in W_i^2} x_j \geq 1 \quad (i \in V) \quad (74)$$

$$\sum_{i \in V} d_i y_i^1 \geq \alpha \sum_{i \in V} d_i \quad (75)$$

$$\sum_{j \in W_i^1} x_j \geq y_i^1 + y_i^2 \quad (i \in V) \quad (76)$$

$$y_i^2 \leq y_i^1 \quad (i \in V) \quad (77)$$

$$\sum_{j \in W} x_j = p \quad (78)$$

$$x_j \leq p_j \quad (j \in W) \quad (79)$$

$$y_i^1, y_i^2 \in \{0, 1\} \quad (i \in V) \quad (80)$$

$$x_j \geq 0 \text{ and integer} \quad (j \in W). \quad (81)$$

Here, the objective function computes the demand covered twice within r_1 time units, constraints (74) mean that all demand is covered within r_2 . The left-hand side of (76) represents

the number of ambulances covering node i within r_1 units, while the right-hand side is equal to 1 if i is covered once within r_1 units, and equal to 2 if it is covered at least twice within r_1 units. The combination of constraints (75) and (76) ensures that a proportion α of the demand is covered within r_1 . Constraints (77) state that node i cannot be covered at least twice if it is not covered at least once. In constraints (79), p_j can be set equal to 2 since an optimal solution using this upper bound always exists.

Gendreau, Laporte and Semet solve the DSM by means of a tabu search procedure in which neighbour solutions are obtained by generating sequences of ambulance moves to an adjacent location, not unlike what is done in ejection chain methods (see, e.g., REGO and ROUCAIROL, 1996). Comparisons with the linear relaxation value of the model indicate that the heuristic typically yields optimal or near-optimal solutions.

6.3 Probabilistic models with extra coverage

None of the models introduced so far takes into account that ambulances are not always available to answer a call. A way around this is to assume that each ambulance has a probability p , called *busy fraction*, of being unavailable. This value is obtained by dividing the total time spent by all ambulances on all calls by the total ambulance time available. If $i \in V$ is covered by k ambulances, then the expected demand covered at that node is $E_{i,k} = d_i(1 - q^k)$ and the marginal contribution of the k^{th} ambulance is $E_{i,k} - E_{i,k-1} = d_i(1 - q)q^{k-1}$.

In the *Maximum Expected Covering Location Model* (MEXCLP) of DASKIN (1983), up to p ambulances may be located in total, and more than one vehicle may be located at the same node. Let y_{ik} be a binary variable equal to 1 if and only if node $i \in V$ is covered by at least k ambulances. The model can be written as follows:

$$\text{(MEXCLP) Maximize} \quad \sum_{i \in V} \sum_{k=1}^p d_i(1 - q)q^{k-1}y_{ik} \quad (82)$$

$$\text{subject to} \quad \sum_{j \in W_i} x_j \geq \sum_{k=1}^p y_{ik} \quad (i \in V) \quad (83)$$

$$\sum_{j \in W} x_j \leq p \quad (84)$$

$$x_j \geq 0 \quad \text{and integer} \quad (j \in W) \quad (85)$$

$$y_{ik} \in \{0, 1\} \quad (i \in V, k = 1, \dots, p). \quad (86)$$

The validity of this model stems from the fact that the objective function is concave in k . Therefore, if $y_{ik} = 1$, then $y_{ih} = 1$ for $h \leq k$. Since the objective is to be maximized, both (83) and (84) will be satisfied as equalities. It follows that the two sides of (83) will be equal to the number of ambulances covering node $i \in V$.

An application of MEXCLP to the City of Bangkok ($|V| = 59$, $|W| = 46$, $10 \leq p \leq 30$) by FUJIWARA et al. (1987) has shown that without reducing expected coverage, the number of ambulances could be reduced to 15 from the current 21. REPEDE and BERNARDO (1994)

have proposed TIMEXCLP, a dynamic implementation of MEXCLP, in which travel speeds are allowed to vary during the day. GOLDBERG et al. (1990) have worked with stochastic travel times. They compute the probability that a given demand point i will be covered based on three probabilities: the probability that an ambulance located at the s^{th} preferred site of i can reach i within eight minutes, the probability that this ambulance is available, and the probability that ambulances located at less preferred sites are not available. Experiments conducted on data collected in Tucson, Arizona, have shown that the use of this model could increase the expected covered demand from 24% to 53.1%.

REVELLE and HOGAN (1989) have developed two chance constrained maximal covering location models, called the *Maximal Availability Location Problem* (MALP I and MALP II), in which the constraint

$$1 - q^{\sum_{j \in W_i} x_j} \geq \alpha \quad (i \in V) \quad (87)$$

ensures that each demand point is covered with probability at least equal to α . This constraint, which can be linearized as

$$\sum_{j \in W_i} x_j \geq \lceil \log(1 - \alpha) / \log q \rceil = b \quad (i \in V), \quad (88)$$

can be used instead of (46) in LSCM. In MALP I, Hogan and ReVelle maximize the total demand covered with b ambulances, subject to a global availability of p ambulances. Defining binary variables y_{jk} as in MEXCLP, the MALP I model can then be written as:

$$\text{(MALP I) Maximize} \quad \sum_{i \in V} d_i y_{ib} \quad (89)$$

$$\text{subject to} \quad \sum_{k=1}^b y_{jk} \leq \sum_{j \in W_i} x_j \quad (i \in V) \quad (90)$$

$$y_{ik} \leq y_{i,k-1} \quad (i \in V, k = 2, \dots, b) \quad (91)$$

$$\sum_{j \in W} x_j = p \quad (92)$$

$$x_j \in \{0, 1\} \quad (j \in W) \quad (93)$$

$$y_{ik} \in \{0, 1\} \quad (i \in V, k = 1, \dots, p). \quad (94)$$

Here, constraints (91) are required since the concavity property observed in MEXCLP no longer holds.

In MALP II, ReVelle and Hogan estimate a busy fraction q_i associated with each i , as the ratio of the total duration of all calls associated to i to the total ambulance time in W_i . The major difficulty with this is that q_i cannot be computed *a priori* because it is an output of the model. An iterative procedure is then required to solve MALP II approximately. Finally we mention the existence of two studies, by BATTA et al. (1989) and by MARIANOV and REVELLE (1994), whose aim is to better approximate the busy fraction of the whole system or associated with a particular location. These are based in part on Larson's (1974) hypercube model.

Finally, BALL and LIN (1993) have developed an extension of LSCM, called *Rel-P*, which incorporates a linear constraint on the number of vehicles required to achieve a given reliability level. The model contains binary variables x_{jk} equal to 1 if and only if k ambulances are located at node $j \in W$, and constants c_{jk} equal to the cost of locating k vehicles at site j . An upper bound p_j is imposed on the number of ambulances located at site j . Their model is as follows:

$$\text{(Rel-P)} \quad \text{Minimize} \quad \sum_{j \in J} \sum_{1 \leq k \leq p_j} c_{jk} x_{jk} \quad (95)$$

$$\text{subject to} \quad \sum_{1 \leq k \leq p_j} x_{jk} \leq 1 \quad (j \in W) \quad (96)$$

$$\sum_{j \in W_i} \sum_{1 \leq k \leq p_j} a_{jk} x_{jk} \geq b_i \quad (i \in V) \quad (97)$$

$$x_{jk} \in \{0, 1\} \quad (j \in W, 1 \leq k \leq p_j). \quad (98)$$

In constraints (97), the constants a_{jk} and b_i are computed to ensure that given the number of ambulances covering demand point i , the probability of being unable to answer a call does not exceed a certain value. The computation of the a_{jk} and b_i coefficients are in fact carried out by using an upper bound on that probability.

Ball and Lin incorporate valid inequalities to their model which is then solved by means of a standard branch-and-bound code for integer linear programming.

6.4 A dynamic model

In practice, ambulances are often relocated over time in order to always ensure an adequate coverage. GENDREAU et al. (2001) have developed a dynamic relocation model where a new redeployment can in principle be implemented at each instant t at which an ambulance is dispatched to a call. The model is based on the DSM developed by the same authors (GENDREAU et al., 1997). It constitutes, to our knowledge, the only available dynamic ambulance relocation tool. In addition to the standard coverage and site capacity constraints, the model takes into account a number of practical considerations inherent to the dynamic nature of the problem: 1) vehicles moved in successive redeployments cannot always be the same; 2) repeated round trips between the same two location sites must be avoided; 3) long trips between the initial and final location sites must be avoided.

The dynamic aspect of the redeployment model is captured by time dependent constants $M_{j\ell}^t$ equal to the cost of repositioning, at time t , ambulance ℓ from its current site to site $j \in W$. This includes the case where site j coincides with the current location of the ambulance, i.e., $M_{j\ell}^t = 0$. The constant $M_{j\ell}^t$ captures some of the history of ambulance ℓ . If it has been moved frequently prior to time t , then $M_{j\ell}^t$ will be larger. If moving ambulance ℓ to site j violates any of the above constraints, then the move is simply disallowed. Binary variables $y_{j\ell}$ are equal to 1 if and only if ambulance ℓ is moved to site j . The *Dynamic Double Standard*

Model at Time t (DDSM^{*t*}) can now be described:

$$(DDSM^t) \quad \text{Maximize} \quad \sum_{i \in V} d_i y_i^2 - \sum_{j \in W} \sum_{\ell=1}^p M_{j\ell}^t x_{j\ell} \quad (99)$$

$$\text{subject to} \quad \sum_{j \in W_i^2} \sum_{\ell=1}^p x_{j\ell} \geq 1 \quad (i \in V) \quad (100)$$

$$\sum_{i \in V} d_i y_i^1 \geq \alpha \sum_{i \in V} d_i \quad (101)$$

$$\sum_{j \in W_i^1} \sum_{\ell=1}^p x_{j\ell} \geq y_i^1 + y_i^2 \quad (i \in V) \quad (102)$$

$$y_i^2 \leq y_i^1 \quad (i \in V) \quad (103)$$

$$\sum_{j \in W} x_{j\ell} = 1 \quad (\ell = 1, \dots, p) \quad (104)$$

$$\sum_{\ell=1}^p x_{\ell j} \leq p_j \quad (j \in W) \quad (105)$$

$$y_i^1, y_i^2 \in \{0, 1\} \quad (i \in V) \quad (106)$$

$$s_{j\ell} \in \{0, 1\} \quad (j \in W, \ell = 1, \dots, p). \quad (107)$$

Apart from variables $x_{j\ell}$, all variables, parameters and constraints of this model can be interpreted as in the static case. The objective function is the demand covered twice within r_1 time units, minus the sum of penalties associated with ambulance relocations at time t .

GENDREAU et al. (2001) solve DDSM^{*t*} by constructing a redeployment table in which the first column gives the list of all ambulances that could possibly be dispatched to the next call, and the second column gives the redeployment plan associated with possible ambulance. Whenever an ambulance is dispatched, the associated relocation plan is implemented and the table is then recomputed from scratch. The authors have used the GENDREAU et al. (1997) tabu search algorithm to compute the relocation strategies and they have also made use of a simple parallel computing strategy: each of 16 processors was assigned the computation of a series of rows of the redeployment table but no communication took place between the processors. The success of this approach rests on the capability of recomputing the entire table between successive calls. Simulations performed on real data from Montreal have shown that this was indeed possible in 95% of all cases. Out of all calls, 38% required at least one ambulance relocation and in only 0.05% of the situations were more than five ambulance relocations necessary. Whenever the table cannot be fully computed between two calls, it is deleted at the next call and no relocation takes place.

Acknowledgement

This work was supported by the Natural Sciences and Engineering Research Council of Canada under grants 227837-00, OGP0039862 and 36662-01. This support is gratefully acknowledged. Most of the material in Section 5 is the result of discussions among the members of a project team designing and implementing scheduling technology for air taxi services, consisting of Mo Bazaraa, Marcos Coycoolea, Daniel Espinoza, Yongpei Guan, Renan Garcia, George Nemhauser, and Martin Savelsbergh, from Georgia Tech and Alex Khmelnitsky and Eugene Taita from Jetson Systems.

References

- T.F. ADAMS. *Police Operations*. Prentice-Hall, Englewood Cliffs, NJ, 1997.
- A.S. ALFA. "Scheduling of Vehicles for Transportation of Elderly." *Transportation Planning and Technology*, **11**:203-212 (1986).
- E. ANGELELLI AND R. MANSINI. "The Vehicle Routing Problem with Time Windows and Simultaneous Pick-up and Delivery." In A. Klose, M.G. Speranza and L.N. Van Wassenhove, editors, *Quantitative Approaches to Distribution Logistics and Supply Chain Management*, Lecture Notes in Economics and Mathematical Systems, pages 249-267. Springer-Verlag, 2001.
- E. BALAS, M. FISCHETTI AND W.R. PULLEYBLANK. "The precedence-constrained asymmetric traveling salesman polytope." *Mathematical Programming*, **68**:241-265 (1995).
- M.O. BALL AND L.F. LIN. "A Reliability Model Applied to Emergency Service Vehicle Location." *Operations Research*, **41**:18-36 (1993).
- R. BATTA, J.M. DOLAN AND N.N. KRISHNAMURTY. "The Maximal Expected Covering Location Problem: Revisited." *Transportation Science*, **23**:277-287 (1989).
- R. BENT AND P. VAN HENTENRYCK. "Scenario-Based Planning for Partially Dynamic Vehicle Routing with Stochastic Customers." *Operations Research* (2004). Forthcoming.
- I. BENYAHIA AND J.-Y. POTVIN. "Decision Support for Vehicle Dispatching using Genetic Programming." *IEEE Transactions on Systems, Man and Cybernetics*, **28**:306-314 (1998).
- D.J. BERTSIMAS AND G.J. VAN RYZIN. "A Stochastic and Dynamic Vehicle Routing Problem in the Euclidean Plane." *Operations Research*, **39**:601-615 (1991).
- L. BIANCO, A. MINGOZZI, S. RICCIARDELLI AND M. SPADONI. "Exact and Heuristic Procedures for the Traveling Salesman Problem with Precedence Constraints, Based on Dynamic Programming." *INFOR*, **32**:19-31 (1994).

- L.D. BODIN AND T. SEXTON. "The Multi-Vehicle Subscriber Dial-a-Ride Problem." *TIMS Studies in Management Science*, **26**:73-86 (1986).
- R. BORNDÖRFER, M. GRÖTSCHER, F. KLOSTERMEIER AND C. KÜTTNER. "Telebus Berlin: Vehicle scheduling in a dial-a-ride system." Technical Report SC 97-23, Konrad-Zuse-Zentrum für Informationstechnik Berlin, 1997.
- L. BROTCORNE, G. LAPORTE AND F. SEMET. "Ambulance Location and Relocation Models." *European Journal of Operational Research*, **147**:451-463 (2003).
- A. CAMPBELL AND M.W.P. SAVELSBERGH. "Incentive Schemes for Attended Home Delivery Services." Technical report, Georgia Institute of Technology, The Logistics Institute, 2003. TLI-03-04.
- A. CAMPBELL AND M.W.P. SAVELSBERGH. "Decision Support for Consumer Direct Grocery Initiatives." *Transportation Science* (2004). Forthcoming.
- N. CHRISTOFIDES. "Worst Case Analysis of a New Heuristic for the Travelling Salesman Problem." Technical Report Report 388, Graduate School of Industrial Administration, Carnegie-Mellon University, Pittsburgh, PA, 1976.
- R.L. CHURCH AND C.S. REVELLE. "The Maximal Covering Location Problem." *Papers of the Regional Science Association*, **32**:101-118 (1974).
- J.-F. CORDEAU. "A Branch-and-Cut Algorithm for the Dial-a-Ride Problem." Technical Report Technical Report CRT-2003-24, Centre for Research on Transportation, 2003.
- J.-F. CORDEAU, M. GENDREAU AND G. LAPORTE. "A tabu search heuristic for periodic and multi-depot vehicle routing problems." *Networks*, **30**:105-119 (1997).
- J.-F. CORDEAU AND G. LAPORTE. "The Dial-a-Ride Problem (DARP): Variants, Modeling Issues and Algorithms." *4OR - Quarterly Journal of the Belgian, French and Italian Operations Research Societies*, **1**:89-101 (2003a).
- J.-F. CORDEAU AND G. LAPORTE. "A Tabu Search Heuristic for the Static Multi-Vehicle Dial-a-Ride Problem." *Transportation Research B*, **37**:579-594 (2003b).
- L. COSLOVICH, R. PESENTI AND W. UKOVICH. "A Two-Phase Insertion Technique of Unexpected Customers for a Dynamic Dial-a-Ride Problem." Technical Report Working paper, Università di Trieste, Italy, 2003.
- F.H. CULLEN, J.J. JARVIS AND H.D. RATLIFF. "Set Partitioning Based Heuristics for Interactive Routing." *Networks*, **11**:125-143 (1981).
- M.S. DASKIN. "A Maximum Expected Location Model: Formulation, Properties and Heuristic Solution." *Transportation Science*, **7**:48-70 (1983).
- M.S. DASKIN AND E.H. STERN. "A Hierarchical Objective Set Covering Model for Emergency Medical Service Vehicle Deployment." *Transportation Science*, **15**:137-152 (1981).

- G. DESAULNIERS, J. DESROSIERS, A. ERDMANN, M.M. SOLOMON AND F. SOUMIS. "VRP with Pickup and Delivery." In P. Toth and D. Vigo, editors, *The Vehicle Routing Problem*, pages 225-242. SIAM Monographs on Discrete Mathematics and Applications, Philadelphia, 2002.
- J. DESROSIERS, Y. DUMAS AND F. SOUMIS. "A Dynamic Programming Solution of the Large-Scale Single-Vehicle Dial-a-Ride Problem with Time Windows." *American Journal of Mathematical and Management Sciences*, 6:301-325 (1986).
- J. DESROSIERS, Y. DUMAS, F. SOUMIS, S. TAILLEFER AND D. VILLENEUVE. "An Algorithm for Mini-Clustering in Handicapped Transport." Technical Report G-91-22, GERAD, HEC Montréal, 1991.
- R.B. DIAL. "Autonomous Dial-A-Ride Transit: Introductory Survey." *Transportation Research C*, 5:261-275 (1995).
- Y. DUMAS, J. DESROSIERS AND F. SOUMIS. "Large Scale Multi-Vehicle Dial-a-Ride Problems." Technical Report G-89-30, GERAD, HEC Montréal, 1989.
- Y. DUMAS, J. DESROSIERS AND F. SOUMIS. "The Pickup and Delivery Problem with Time Windows." *European Journal of Operational Research*, 54:7-22 (1991).
- Y. DUMAS, F. SOUMIS AND J. DESROSIERS. "Optimizing the Schedule for a Fixed Vehicle Path with Convex Inconvenience Costs." *Transportation Science*, 24:145-152 (1990).
- M. FISCHETTI AND P. TOTH. "An Additive Bounding Procedure for Combinatorial Optimization Problems." *Operations Research*, 37:319-328 (1989).
- L. FU. "Scheduling dial-a-ride paratransit under time-varying, stochastic congestion." *Transportation Research B*, 36:485-506 (2002).
- O. FUJIWARA, T. MAKJAMROEN AND K.K. GUPTA. "Ambulance Deployment Analysis: A Case Study of Bangkok." *European Journal of Operational Research*, 31:9-18 (1987).
- M. R. GAREY AND D. S. JOHNSON. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. Freeman, San Francisco, 1979.
- M. GENDREAU, F. GUERTIN, J.-Y. POTVIN AND R. SÉGUIN. "Neighborhood Search Heuristics for a Dynamic Vehicle Dispatching Problem with Pick-ups and Deliveries." Technical Report CRT-98-10, Centre de recherche sur les transports, Montreal, 1998.
- M. GENDREAU, G. LAPORTE AND F. SEMET. "Solving an Ambulance Location Model by Tabu Search." *Location Science*, 5:75-88 (1997).
- M. GENDREAU, G. LAPORTE AND F. SEMET. "A Dynamic Model and Parallel Tabu Search Heuristic for Real-Time Ambulance Relocation." *Parallel Computing*, 27:1641-1653 (2001).

- M. GENDREAU, G. LAPORTE AND D. VIGO. "Heuristics for the Traveling Salesman Problem with Pickup and Delivery." *Computers & Operations Research*, **26**:699-714 (1999).
- F. GLOVER. "Ejection Chains, Reference Structures and Alternating Path Methods for Traveling Salesman Problems." *Discrete Applied Mathematics*, **65**:223-253 (1996).
- F. GLOVER AND M. LAGUNA. *Tabu Search*. Kluwer, Boston, 1997.
- J. GOLDBERG, R. DIETRICH, J.M. CHEN AND M.G. MITWASI. "Validating and Applying a Model for Locating Emergency Medical Services in Tucson, AZ." *European Journal of Operational Research*, **49**:308-324 (1990).
- M. GRÖTSCHEL AND M.W. PADBERG. "Polyhedral Theory." In E.L. Lawler, J.K. Lenstra, A.H.G. Rinnooy Kan and D.B. Shmoys, editors, *The Traveling Salesman Problem*, pages 251-305. Wiley, New York, 1985.
- P. HEALY AND R. MOLL. "A New Extension of Local Search Applied to the Dial-a-Ride Problem." *European Journal of Operational Research*, **83**:83-104 (1995).
- H. HERNÁNDEZ-PÉREZ AND J.-J. SALAZAR-GONZÁLEZ. "A Branch-and-Cut Algorithm for the Traveling Salesman Problem with Pickups and Deliveries." *Discrete Applied Mathematics* (2004). Forthcoming.
- K. HOGAN AND C.S. REVELLE. "Concepts and Applications of Backup Coverage." *Management Science*, **34**:1434-1444 (1986).
- J.H. HOLLAND. *Adaptation in Natural and Artificial Systems*. The MIT Press, 1992.
- M.E.T. HORN. "Fleet Scheduling and Dispatching for Demand-Responsive Passenger Services." *Transportation Research C*, **10**:35-63 (2002).
- B. HUNSAKER AND M. SAVELSBERGH. "Efficient testing for dial-a-ride problems." *Operations Research Letters*, **30**:169-173 (2002).
- I. IOACHIM, J. DESROSIERS, Y. DUMAS AND M.M. SOLOMON. "A request clustering algorithm for door-to-door handicapped transportation." *Transportation Science*, **29**:63-78 (1995).
- J. JAW, A.R. ODONI, H.N. PSARAFTIS AND N.H.M. WILSON. "A Heuristic Algorithm for the Multi-Vehicle Advance-Request Dial-a-Ride Problem with Time Windows." *Transportation Research B*, **20**:243-257 (1986).
- B. KALANTARI, A.V. HILL AND S.R. ARORA. "An Algorithm for the Traveling Salesman Problem with Pickup and Delivery Customers." *European Journal of Operational Research*, **22**:377-386 (1985).
- J.R. KOZA. *Genetic Programming*. The MIT Press, Cambridge, MA, 1992.
- R.C. LARSON. *Urban Police Patrol Analysis*. The MIT Press, Cambridge, MA, 1972.

- H.C. LAU AND Z. LIANG. "Pickup and Delivery with Time Windows: Algorithms and Test Case Generation." *International Journal on Artificial Intelligence Tools*, **11**:455-472 (2002).
- F. LECLERC AND J.-Y. POTVIN. "Genetic Algorithms for Vehicle Dispatching." *International Transactions in Operational Research*, **4**:391-400 (1997).
- S. LIN. "Computer Solutions of the Travelling Salesman Problem." *Bell System Technical Journal*, **44**:2245-2269 (1965).
- S. LIN AND B. W. KERNIGHAN. "An Effective Heuristic Algorithm for the Traveling-Salesman Problem." *Operations Research*, **21**:498-516 (1973).
- J.D.C. LITTLE, K.G. MURTY, D.W. SWEENEY AND C. KAREL. "An Algorithm for the Traveling Salesman Problem." *Operations Research*, **11**:972-989 (1963).
- O. B. G. MADSEN, K. TOSTI AND J. VAELDS. "A Heuristic Method for Dispatching Repair Men." *Annals of Operations Research*, **61**:213-226 (1995).
- M.B. MANDELL. "Covering Models for Two-Tiered Emergency Medical Services Systems." *Location Science*, **6**:355-368 (1998).
- V. MARIANOV AND C.S. REVELLE. "The Queueing Probabilistic Location Set Covering Problem and Some Extensions." *Socio-Economic Planning Sciences*, **28**:167-178 (1994).
- V. MARIANOV AND C.S. REVELLE. "Siting Emergency Services." In Z. Drezner, editor, *Facility Location*, pages 199-223. Springer, New York, 1995.
- S. MITROVIĆ-MINIĆ. *The Dynamic Pick-up and Delivery Problem with Time Windows*. Ph.D. thesis, School of Computing Science, Simon Fraser University, Canada, 2001.
- S. MITROVIĆ-MINIĆ, R. KRISHNAMURTI AND G. LAPORTE. "Double-Horizon Based Heuristics for the Dynamic Pickup and Delivery Problem with Time Windows." *Transportation Research B*, **38**:669-685 (2004).
- S. MITROVIĆ-MINIĆ AND G. LAPORTE. "Waiting Strategies for the Dynamic Pickup and Delivery Problem with Time Windows." *Transportation Research B*, **38**:635-655 (2004).
- W.P. NANRY AND J.W. BARNES. "Solving the pickup and delivery problem with time windows using reactive tabu search." *Transportation Research B*, **34**:107-121 (2000).
- H.N. PSARAFTIS. "A Dynamic Programming Approach to the Single-Vehicle, Many-to-Many Immediate Request Dial-a-Ride Problem." *Transportation Science*, **14**:130-154 (1980).
- H.N. PSARAFTIS. "An Exact Algorithm for the Single-Vehicle Many-to-Many Dial-a-Ride Problem with Time Windows." *Transportation Science*, **17**:351-357 (1983).

- H.N. PSARAFTIS. "Dynamic Vehicle Routing Problems." In B.L. Golden and A.A. Assad, editors, *Vehicle Routing: Methods and Studies*, pages 223-248. North-Holland, Amsterdam, 1988.
- C. REGO AND C. ROUCAIROL. "A Parallel Tabu Search Algorithm Using Ejection Chains for the Vehicle Routing Problem." In I.H. Osman and J.P. Kelly, editors, *Meta-Heuristics: Theory and Applications*, pages 661-675. Kluwer, Boston, 1996.
- J.F. REPEDE AND J.J. BERNARDO. "Developing and Validating a Decision Support System for Locating Emergency Medical Vehicles in Louisville, Kentucky." *European Journal of Operational Research*, **75**:567-581 (1994).
- C.S. REVELLE AND K. HOGAN. "The Maximum Availability Location Problem." *Transportation Science*, **23**:192-200 (1989).
- Y. ROCHAT AND É. D. TAILLARD. "Probabilistic diversification and intensification in local search for vehicle routing." *Journal of Heuristics*, **1**:147-167 (1995).
- K.S. RULAND. *Polyhedral Solution to the Pickup and Delivery Problem*. Ph.D. thesis, Sever Institute of Technology, Washington University, 1995.
- K.S. RULAND AND E.Y. RODIN. "The Pickup and Delivery Problem: Faces and Branch-and-Cut Algorithm." *Computers and Mathematics with Applications*, **33**:1-13 (1997).
- D.E. RUMELHART, G.E. HINTON AND R.J. WILLIAMS. "Learning Internal Representations by Error Propagation." In D.E. Rumelhart and J.L. McClelland, editors, *Parallel Distributed Processing, Vol. 1*, pages 318-364. The MIT Press, Cambridge, MA, 1986.
- M.W.P. SAVELSBERGH. "Local Search in Routing Problems with Time Windows." *Annals of Operations Research*, **4**:285-305 (1985).
- M.W.P. SAVELSBERGH. "En Efficient Implementation of Local Search Algorithms for Constrained Routing Problems." *European Journal of Operational Research*, **47**:75-85 (1990).
- M.W.P. SAVELSBERGH. "The Vehicle Routing Problem with Time Windows: Minimizing Route Duration." *ORSA Journal on Computing*, **4**:146-154 (1992).
- M.W.P. SAVELSBERGH AND M. SOL. "The General Pickup and Delivery Problem." *Transportation Science*, **29**:17-29 (1995).
- M.W.P. SAVELSBERGH AND M. SOL. "DRIVE: Dynamic Routing of Independent Vehicles." *Operations Research*, **46**:474-490 (1998).
- D.A. SCHILLING, D.J. ELZINGA, J. COHON, R.L. CHURCH AND C.S. REVELLE. "The TEAM FLEET Models for Simultaneous Facility and Equipment Siting." *Transportation Science*, **13**:163-175 (1979).

- T. SEXTON AND L.D. BODIN. "Optimizing single vehicle many-to-many operations with desired delivery times: I. Scheduling." *Transportation Science*, **19**:378-410 (1985a).
- T. SEXTON AND L.D. BODIN. "Optimizing single vehicle many-to-many operations with desired delivery times: II. Routing." *Transportation Science*, **19**:411-435 (1985b).
- Y. SHEN, J.-Y. POTVIN, J.-M. ROUSSEAU AND S. ROY. "A Computer Assistant for Vehicle Dispatching with Learning Capabilities." *Annals of Operations Research*, **61**:189-211 (1995).
- D.M. STEIN. "Scheduling Dial-a-Ride Transportation Systems." *Transportation Science*, **12**:232-249 (1978).
- A.J. SWERSEY. "The Deployment of Police, Fire, and Emergency Medical Units." In S.M. Pollock, M.H. Rothkopf and A. Barnett, editors, *Operations Research and the Public Sector, Handbooks in Operations Research and Management Science, Vol. 6*, pages 151-200. North-Holland, 1994.
- M.R. SWIHART AND J.D. PAPASTRAVOU. "A Stochastic and Dynamic Model for the Single-Vehicle Pick-up and Delivery Problem." *European Journal of Operational Research*, **114**:447-464 (1999).
- D. TEODOROVIC AND G. RADIVOJEVIC. "A Fuzzy Logic Approach to Dynamic Dial-A-Ride Problem." *Fuzzy Sets and Systems*, **116**:23-33 (2000).
- C.R. TOREGAS, R. SWAIN, C.S. REVELLE AND L. BERGMAN. "The Location of Emergency Service Facilities." *Operations Research*, **19**:1363-1373 (1971).
- P. TOTH AND D. VIGO. "Fast Local Search Algorithms for the Handicapped Persons Transportation Problem." In I.H. Osman and J.P. Kelly, editors, *Meta-Heuristics: Theory and Applications*, pages 677-690. Kluwer, Boston, 1996.
- P. TOTH AND D. VIGO. "Heuristic Algorithms for the Handicapped Persons Transportation Problem." *Transportation Science*, **31**:60-71 (1997).
- L.J.J. VAN DER BRUGGEN, J.K. LENSTRA AND P.C. SCHUUR. "Variable-Depth Search for the Single-Vehicle Pickup and Delivery problem with Time Windows." *Transportation Science*, **27**:298-311 (1993).
- W.E. WALKER, J.M. CHAIKEN AND E.J. IGNALL. *Fire Department Deployment Analysis*. North-Holland, New York, 1979.
- H. XU, Z.-L. CHEN, S. RAJAGOPAL AND S. ARUNAPURAM. "Solving a Practical Pickup and Delivery Problem." *Transportation Science*, **37**:347-364 (2003).